A Visually Plausible Grasping System for Object Manipulation and Interaction in Virtual Reality Environments

Sergiu Oprea, Pablo Martinez-Gonzalez, Alberto Garcia-Garcia, John Alejandro Castro-Vargas, Sergio Orts-Escolano, and Jose Garcia-Rodriguez

Abstract—Interaction in virtual reality (VR) environments is essential to achieve a pleasant and immersive experience. Most of the currently existing VR applications, lack of robust object grasping and manipulation, which are the cornerstone of interactive systems. Therefore, we propose a realistic, flexible and robust grasping system that enables rich and real-time interactions in virtual environments. It is visually realistic because it is completely user-controlled, flexible because it can be used for different hand configurations, and robust because it allows the manipulation of objects regardless their geometry, i.e. hand is automatically fitted to the object shape. In order to validate our proposal, an exhaustive qualitative and quantitative performance analysis has been carried out. On the one hand, qualitative evaluation was used in the assessment of the abstract aspects such as: hand movement realism, interaction realism and motor control. On the other hand, for the quantitative evaluation a novel error metric has been proposed to visually analyze the performed grips. This metric is based on the computation of the distance from the finger phalanges to the nearest contact point on the object surface. These contact points can be used with different application purposes, mainly in the field of robotics. As a conclusion, system evaluation reports a similar performance between users with previous experience in virtual reality applications and inexperienced users, referring to a steep learning curve.

Index Terms—Virtual Reality, Visualization, Hand Interaction.

1 Introduction

With the advent of affordable VR headsets such as Oculus VR/Go and HTC Vive, many works and projects are using virtual environments for different purposes. Most of VR applications are related to the entertainment industry (i.e. games and 3D cinema) or architectural visualizations, where virtual scene realism is a cornerstone. Currently existing VR systems are limited by their resolution, field-of-view, frame rate, and interaction among other technical specifications. In order to enhance user VR experience, developers are also focused on implementing rich interactions with the virtual environment, allowing the user to explore, interact and manipulate scene objects as in the real world.

Interaction is a crucial feature for training/simulation applications (e.g. flight, driving and medical simulators), and also teleoperation (e.g. robotics), where the user ability to interact and explore the simulated environments is paramount for achieving an immersive experience. For this purpose, most of VR devices come with a pair of hand-held controllers which are fully tracked in 3D space and specifically designed for interaction. One of the most basic interaction tasks is object grasping and manipulation. In order to achieve an enjoyable experience in VR, a realistic, flexible and real-time grasping system is needed. However, grasp synthesis in manipulation tasks is not straightforward because of the unlimited number of different hand configurations, the variety of object types and their geometries, and also due to the selection of the most suitable grasp for every different object in terms of realism, kinematics and physics.

Currently existing real-time approaches in VR are purely animation-driven, completely relying on the animations realism. Moreover, these approaches are constrained to a limited number of simple object geometries and unable to deal with unknown objects. For every different object type and geometry, predefined animations are needed. This fact hinders the user experience, limiting its interaction capabilities. For a complete immersion user should be able to interact and manipulate different virtual objects as in the real world.

In this paper, we propose a real-time grasping system for object interaction in virtual reality environments. We aim to achieve natural and visually plausible interactions in photo-realistic environments rendered by Unreal Engine. Taking advantage of headset tracking and motion controllers, a human operator can be embodied in such environments as a virtual human or robot agent to freely navigate and interact with objects. Our grasping system is able to deal with different object geometries, without the need of a predefined grasp animation for each. With our approach, fingers are automatically fitted to object shape and geometry. We constrain hand finger phalanges motion checking in real-time for collisions with the object geometry.

Our grasping system was analyzed both qualitatively and quantitatively. On one side, for the qualitative analysis, grasping system was implemented in a photorealistic envi-

• *Sergiu Oprea, Pablo M. Gonzalez, Alberto G. Garcia, John Alejandro C. Vargas, Sergio O. Escolano, and Jose G. Garcia are with the 3D Perception Lab at the University of Alicante, Spain.*
E-mails: soprea@dtic.ua.es, pmartinez@dtic.ua.es, agarcia@dtic.ua.es, ja-castro@dtic.ua.es, sorts@ua.es, jgarcia@dtic.ua.es.

Manuscript received April 19, 2005; revised August 26, 2015.

Fig. 1: Examples of interaction with objects extracted from the YCB dataset [1] in a photorealistic virtual environment, where the user: grabs a pear and mustard (top left), serves wine in a glass (top right), cooks mushrooms in a frying pan with a little oil (bottom left), and uses the mortar (bottom right).

ronment where the user is freely able to interact with real world objects extracted from the YCB dataset [1] (see Figure 1). The qualitative evaluation is based on a questionnaire that will address the user interaction experience in terms of realism during object manipulation and interaction, system flexibility and usability, and general VR experience. On the other side, a quantitative grasping system analysis was carried out, contrasting the elapsed time a user needs in grasping an object and grasp quality based on a novel error metric which quantifies the overlapping between hand fingers and grasped object.

From the quantitative evaluation, we obtain individual errors for the last two phalanges of each finger, the time user needed to grasp the object and also the contact points. This information alongside other provided by UnrealROX [2] such as depth mpas, instance segmentations, normal maps, 3D bounding boxes and 6D object pose (see Figure 8), enables different robotic applications as described in Section 6.

In summary, we make the three following contributions:

- We propose a real-time, realistic looking and flexible grasping system for natural interaction with arbitrary shaped objects in virtual reality environments;
- We propose a novel metric and procedure to analyze visual grasp quality in VR interactions quantifying hand-object overlapping;
- We provide the contact points extracted during the interaction in both local and global system coordinates.

The rest of the paper is structured as follows. First of all, Section 2 analyzes the latest works related to object interaction and manipulation in virtual environments. The core of this work is comprised in Section 3 where our approach is described in detail. Then, the performance analysis, with the qualitative and our novel quantitative evaluations, is discussed in Section 4. Analysis results are reported in Section 5. Then, several applications are discussed in Section 6. After that, limitations of our approach are covered in Section 7 alongside several feature works. Finally, some conclusions are drawn in the last Section 8.

2 Related works

Computer graphics are fundamental for virtual reality applications, bringing realistic environments to users. However, currently existing virtual reality systems come with a pair of handheld devices specifically designed to enable user interaction with the virtual environment. This has led researchers to focus on designing efficient and realistic interactions with the virtual environment in order to improve the user experience. In spite of existing approaches, VR interaction remains an open problem for researchers and companies.

Grasping action is the most basic component of any interaction and it is composed of three major components [3]. The first one is related to the process of approaching the arm and hand to the target object, considering the overall body movement. The second component focuses on the hand and body pre-shaping before the grasping action. Finally, the last component fits the hand to the geometry of the object by closing each of the fingers until contact is established.

Our work and most of the currently existing works about virtual reality interaction and grasping, relate to the third component of the grasping action. Most of the currently existing approaches in solving this problem are data-driven. This is using predefined animations for concrete object geometries which are stored in a database. The keys of data-driven approaches are to effectively index large datasets in order to quickly match unknown object geometries with existing hand poses in the database.

In this section we will only mention the most recent works and approaches which are mostly aligned with our proposal and preferably related to virtual reality applications in which interaction is done by means of handheld devices, such as Oculus Touch controllers. Moreover and for an in-depth insight, a detailed review of the advances conducted in hand modeling and animation was published by Wheatland et al. [4]. At the same time and regarding 3D object selection techniques in virtual environments, a review was published by Argelaguet et al. [5].

2.1 Data-driven approaches

Grasping data-driven approaches have existed since a long time ago [3]. These methods are based on large databases of predefined hand poses selected using user criteria or based on grasp taxonomies (i.e. final grasp poses when an object was successfully grasped) which provide us the ability to discriminate between different grasp types.

From this database, grasp poses are selected according with given object shape and geometry [6] [7]. Li et al. [6] construct a database with different hand poses and also object shapes and sizes. Despite having a good database, the process of hand poses selection is not straightforward since there can be multiple equally valid possibilities for the same gesture. To address this problem, Li et al. [6] proposed a shape-matching algorithm which returns multiple potential grasp poses.

The selection process is also constrained by the hand high degree of freedom (DOF). In order to deal with dimensionality and redundancy many researchers have used techniques such as principal component analysis (PCA) [8] [9]. For the same purpose, Jorg et al. [10] studied the correlations between hand DOFs aiming to simplify hand models reducing DOF number. The results suggest to simplify hand models by reducing DOFs from 50 to 15 for both hands in conjunction without loosing relevant features.

2.2 Hybrid data-driven approaches

In order to achieve realistic object interactions, physical simulations on the objects should also be considered [11] [12] [13]. Moreover, hand and finger movement trajectories need to be both, kinematically and dynamically valid [14]. Pollard et al. [11] simulate hand interaction, such as two hands grasping each other in the handshake gesture. Bai et al. [13] simulate grasping an object, drop it on a specific spot on the palm and let it roll on the hand palm. A limitation of this approach is that information about the object must be known in advance, which disable robot to interact with unknown objects. By using an initial grasp pose and a desired object trajectory, the algorithm proposed by Liu [15] can generate physically-based hand manipulation poses varying the contact points with the object, grasping forces and also joint configurations. This approach works well for complex manipulations such as twist-opening a bottle. Ye and Liu [14] reconstruct a realistic hand motion and grasping generating feasible contact point trajectories. Selection of valid motions is defined as a randomized depth-first tree traversal, where nodes are recursively expanded if they are kinematically and dynamically feasible. Otherwise, backtracking is performed in order to explore other possibilities.

2.3 Virtual reality approaches

This section is limited to virtual reality interaction using VR motion controllers, avoiding glove-based and bare-hand approaches. Implementation of the aforementioned techniques in virtual reality environments is a difficult task cause optimizations are needed to keep processes running in real time. Most of current existing approaches for flexible and realistic grasping are not suitable for real-time interaction. VR developers aim to create fast solutions with realistic and natural interactions.

Recent approaches are directly related to the entertainment industry, i.e. video games. An excellent example is *Lone Echo*, a narrative adventure game which consists of manipulating tools and objects for solving puzzles. Hand animations are mostly procedurally generated, enabling grasping of complex geometries regardless their grasp angle. This approach [16] is based on a graph traversal heuristic which searches intersections between hand fingers and object surface mesh triangles. A* heuristic find the intersection that is nearest to the palm and also avoid invalid intersections. After calculating angles to make contact with each intersection point, highest angle is selected and fingers are rotated accordingly.

Mostly implemented solutions in VR are animation-based [17] [18] [19]. These approaches are constrained to a limited number of simple object geometries and are unable to deal with unknown objects. Movements are predefined for concrete object geometries, hindering user interaction capabilities in the virtual environment. In [17], distance grab selection technique is implemented to enhance the user comfort when interacting in small play areas, while sitting or for grabbing objects on the floor. Grasping system is based on three trigger volumes attached to each hand: two small cylinders for short-range grasp, and a cone for long-range grabbing. Based on this approach, we used trigger volumes attached to finger phalanges to control its movement and detect object collisions more precisely. In this way we achieve a more flexible and visually plausible grasping system enhancing immersion and realism during interactions.

3 Grasping System

With the latest advances in rendering techniques, visualization of virtual reality (VR) environments is increasingly more photorealistic. Besides graphics, which are the cornerstone of most VR solutions, interaction is also an essential part to enhance the user experience and immersion. VR scene content is portrayed in a physically tangible way, inviting users to explore the environment, and interact or manipulate represented objects as in the real world. VR devices aim to provide very congruent means of primary interaction, described as a pair of handheld devices with very accurate 6D one-to-one tracking. The main purpose is to create rich interactions producing memorable and satisfying VR experiences.

Most of the currently available VR solutions and games lack of a robust and natural object manipulation and interaction capabilities. This is because, bringing natural and intuitive interactions to VR is not straightforward, which makes VR development challenging at this stage. Interactions need

to be in real-time and maintaining a high and solid frame rate, directly mapping user movement to VR input in order to avoid VR sickness (visual and vestibular mismatch). Maintaining the desired 90 frames per second (FPS) in a photorealistic scene alongside complex interactions is not straightforward. This indicates the need of a flexible grasping system designed to naturally and intuitively manipulate unknown objects of different geometries in real-time.

3.1 Overview

Our grasping approach was designed for real-time interaction and manipulation in virtual reality environments by providing a simple, modular, flexible, robust, and visually realistic grasping system. Its main features are described as follows:

- Simple and modular: it can be easily integrated with other hand configurations. Its design is modular and adaptable to different hand skeletals and models.
- Flexible: most of the currently available VR grasp solutions are purely animation-driven, thus limited to known geometries and unable to deal with previously unseen objects. In contrast, our grasping system is flexible as it allows interaction with unknown objects. In this way, the user can freely decide the object to interact with, without any restrictions.
- Robust: unknown objects can have different geometries. However, our approach is able to adapt the virtual hand to objects, regardless of their shape.
- Visually realistic: grasping action is fully controlled by the user, taking advantage of its previous experience and knowledge in grasping daily common realistic objects such as cans, cereal boxes, fruits, tools, etc. This makes resulting grasping visually realistic and natural just as a human would in real life.

The combination of the above described features makes VR interaction a pleasant user experience, where object manipulation is smooth and intuitive.

Our grasping works by detecting collisions with objects through the use of trigger actors placed experimentally on the finger phalanges. A trigger actor is a component from Unreal Engine 4 used for casting an event in response to an interaction, e.g. collision with another object. These components can be of different shapes, such as capsule, box, sphere, etc. In the Figure 2 capsule triggers are represented in green and palm sphere trigger in red. We experimentally placed two capsule triggers on the last two phalanges of each finger. We noticed that this configuration is the most effective in detecting objects collisions. Notice that collision detection is performed for each frame, so, for heavy configurations with many triggers, performance would be harmed.

3.2 Components

Our grasping system is composed of the components represented in the Figure 3. These components are defined as follows:

- Object selection: selects the nearest object to the hand palm. Detection area is determined by the sphere trigger placed on the hand palm (red colored in Figure 2). The sphere trigger returns the world location of all the overlapped actors. As a result, the nearest actor can be determined by computing the distance from each overlapped actor to the center of the sphere trigger. Smallest distance will determine the nearest object, saving its reference for the other components.
- Interaction manager: manages capsule triggers which are attached to finger phalanges as represented in Figure 2. If a capsule trigger reports an overlap event, the movement of its corresponding phalanx is blocked until hand is reopened or the overlapping with the manipulated object is over. The phalanx state (blocked or in movement) will be used as input to the grasping logic component. A phalanx is blocked if there is an overlap of the its corresponding capsule trigger with the manipulated object.
- Finger movement: this component determines the movement of the fingers during the hand closing and opening animations. It ensures a smooth animation avoiding unexpected and unrealistic behavior in finger movement caused neither by a performance drop or other interaction issues. Basically, it monitors each variation in the rotation value of the phalanx. In the case of detecting an unexpected variation (i.e. big variation) during a frame change, missing intermediate values will be interpolated so as to keep finger movement smooth.
- Grasping logic: this component manages when to grab or release an object. This decision is made based on the currently blocked phalanges determined with the interaction manager component. The object is grasped or released based on the following function:

$$f(x) = \begin{cases} true, & \text{if } (th_{ph} \vee palm) \wedge (in_{ph} \vee mi_{ph}) \\ false, & \text{otherwise} \end{cases} \quad (1)$$

, where $x = (th_{ph}, in_{ph}, mi_{ph}, palm)$ is defined as

$$\begin{aligned} th_{ph} &= thumb_{mid} \vee thumb_{dist} \\ in_{ph} &= index_{mid} \vee index_{dist} \\ mi_{ph} &= middle_{mid} \vee middle_{dist} \end{aligned} \quad (2)$$

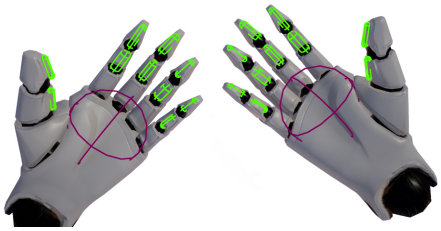


Fig. 2: In green, capsule triggers of the middle and distal phalanges. In purple, sphere trigger of the palm.

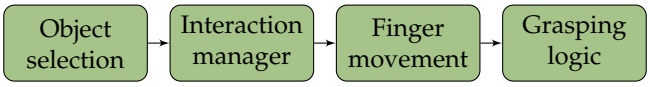


Fig. 3: Pipeline with grasping system components

Equation 1 determines when an object is grasped or released based on the inputs determined in Equation 2 where th_{ph}, in_{ph}, and mi_{ph}, are the thumb, index and middle phalanges respectively. According to human hand morphology, mid and $dist$ subscripts refer to the middle and distal phalanx (e.g. $thumb_{dist}$ references the distal phalanx of thumb finger and at the implementation level it is a boolean value).

3.3 Implementation details

Grasping system has been originally implemented in Unreal Engine 4 (UE4), however, it can be easily implemented in other engines such as Unity, which would also provide us with the necessary components for replicating the system (e.g. overlapping triggers). The implementation consists of UE4 blueprints and has been correctly structured in the components depicted in Figure 3 and described in the previous section. Implementation is available at Github[1].

4 Performance analysis

In order to validate our proposal, a complete performance analysis has been carried out. This analysis covers from a qualitative evaluation, which is prevalent in the assessment of VR systems, to a novel quantitative evaluation. Evaluation methods are briefly described as follows:

- Qualitative evaluation: based on the user experience interacting with real objects from the YCB dataset in a photorealistic indoor scenario. Its purpose is to assess interaction realism, immersion, hand movement naturalness and other qualitative aspects described in Table 1 from the Subsection 4.1, which addresses qualitative evaluation in detail.
- Quantitative evaluation: based on the grasping quality in terms of realism (i.e. how much it is visually plausible). We consider a visually plausible grasp when hand palm or fingers are level with the object surface, as in a real life grasping. However, when dealing with complex meshes, the collision detection precision can be significantly influenced. In this case, fingers could penetrate the object surface, or remain above its surface when a collision was detected earlier than expected. This would result in an unnatural and unrealistic grasp. To visually quantify grasping quality, we purpose a novel error metric based on computing the distance from each capsule trigger to the nearest contact point on the object surface. Quantitative evaluation and the proposed error metric are addressed in detail in Subsection 4.2.

4.1 Qualitative evaluation

Most VR experiments include qualitative and quantitative studies to measure its realism and immersion. Arguably, questionnaires are the default method to qualitatively assess any experience and the vast majority of works include them in one way or another [20] [21] [22]. However, one of the main problems with them is the absence of a standardized set of questions for different experiences that allows for fair and easy comparisons. The different nature of the VR systems and experiences makes it challenging to find a set of evaluation questions that fits them all. Following the efforts of [23] towards a standardized embodiment questionnaire, we analyzed several works in the literature [24] [25] that included questionnaires to assess VR experiences to devise a standard one for virtual grasping systems. Inspired by such works, we have identified three main types of questions or aspects:

- Motor Control: this aspect considers the movement of the virtual hands as a whole and its responsiveness to the virtual reality controllers. Hands should move naturally and their movements must be caused exactly by the controllers without unwanted movements and without limiting or restricting real movements to adapt to the virtual ones.
- Finger Movement: this aspect takes the specific finger movement into account. Such movements must be natural and plausibly. Moreover, they must react properly to the user's intent.
- Interaction Realism: this aspect is related to the interaction of the hand and fingers with objects.

The questionnaire, shown in Table 1, is composed of fourteen questions related to the previously described aspects. Following [23], the users of the study will be presented with such questions right after the end of the experience in a randomized order to limit context effects. In addition, questions must be answered following the 7-point Likert-scale: (+3) strongly agree, (+2) agree, (+1) somewhat agree, (0) neutral, (-1) somewhat disagree, (-2) disagree, and (-3) strongly disagree. Results will be presented as a single embodiment score using the following equations:

ID	Question
	Aspect 1: Motor Control
Q1	I felt like I could control the virtual hands as if it were my own hands
Q2	The movements of the virtual hands were caused by my movements
Q3	I felt as if the movements of the virtual hands were influencing my own movements
Q4	I felt as if the virtual hands were moving by themselves
	Aspect 2: Finger Movement Realism
Q5	It seemed that finger movements were smooth and plausible
Q6	I felt fingers open and close in a natural way
Q7	Fingers react adequately to my intentions
	Aspect 3: Interaction Realism
Q8	I felt like I could grab objects wherever I wanted to
Q9	It seemed as if the virtual fingers were mine when grabbing an object
Q10	I felt that grabbing objects was clumsy and hard to achieve
Q11	It seemed as if finger movement were guided and unnatural
Q12	I felt that grasps were visually correct and natural
Q13	I felt that grasps were physically correct and natural
Q14	It seemed that fingers were adapting properly to the different geometries

TABLE 1: User evaluation questionnaire.

$$\begin{aligned} \text{Motor Control} &= ((Q1 + Q2) - (Q3 + Q4))/4 \\ \text{Finger Movement Realism} &= (Q5 + Q6 + Q7)/3 \\ \text{Interaction Realism} &= ((Q8 + Q9) - (Q10 + Q11) \\ &\quad + Q12 + Q13 + Q14)/7 \end{aligned} \tag{3}$$

, using the results of each individual aspect, we obtain the total embodiment score as follows:

$$\begin{aligned} \text{Score} = (&\text{Motor Control} + \text{Finger Movement Realism} \\ &+ \text{Interaction Realism} * 2)/4 \end{aligned} \tag{4}$$

The interaction realism is the key aspect of this qualitative evaluation. So that, in the Equation 4 we emphasize this aspect by weighting it higher.

1. https://github.com/3dperceptionlab/unrealgrasp

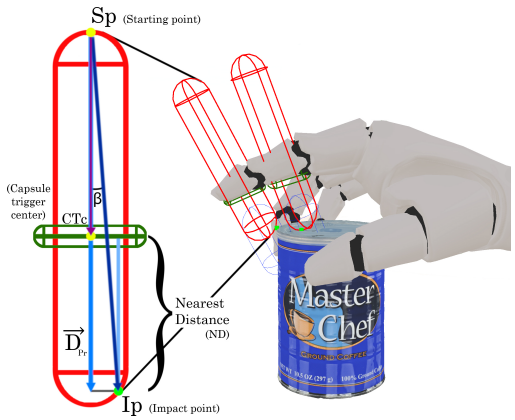


Fig. 4: Distance computation for index finger.

4.2 Quantitative evaluation

With the quantitative evaluation, we aim to evaluate grasping quality in terms of how much it is visually plausible or realistic. In other words, our purpose is to visually quantify our grasping performance, analyzing each finger position and how it fits the object mesh. When a collision is detected by a capsule trigger, we proceed with the calculation of the nearest distance between the finger phalanx surface (delimited by the capsule trigger) and the object mesh (see Equation 8).

In Figure 4 the red capsules are representing 3D sphere tracing volumes which provide information of the nearest collision from the trace starting point to the first contact point on the object surface which intersects the sphere volume. For each finger phalanx with an attached capsule trigger represented in green, we throw a sphere trace obtaining the nearest contact points on the object surface represented as lime colored dots (impact point, Ip). In this representation, the total error for the index finger would be the average of the sum of the distances in millimeters between the surface of each phalanx and the nearest contact point on the object surface (see Equation 9). The nearest distance computation is approximated by an equation that was developed to find the distance between the impact point, and the plane that contains the capsule trigger center point and is perpendicular to the longitudinal axis of the red capsule. Capsule triggers centers are located on the surface of the hand mesh, so this computation should approximate the nearest distance to the mesh well enough, without being computationally too demanding. To compute this distance, we define the following vectors from the three input points (the starting point of the red capsule, the impact point and the capsule trigger center point):

$$\begin{aligned} \overrightarrow{D_{Ip}} &= Ip - Sp \\ \overrightarrow{D_{CTc}} &= CTc - Sp \end{aligned} \tag{5}$$

where $\overrightarrow{D_{Ip}}$ is the vector from the starting point to the impact point, and $\overrightarrow{D_{CTc}}$ vector represents the direction of the longitudinal axis of the red capsule. They are represented in navy blue and purple respectively in Figure 4. Then, we find the cosine of the angle they form through their dot product:

$$\begin{aligned} \overrightarrow{D_{Ip}} \cdot \overrightarrow{D_{CTc}} &= |\overrightarrow{D_{Ip}}| * |\overrightarrow{D_{CTc}}| * cos(\beta) \\ cos(\beta) &= \frac{\overrightarrow{D_{Ip}} \cdot \overrightarrow{D_{CTc}}}{|\overrightarrow{D_{Ip}}| * |\overrightarrow{D_{CTc}}|} \end{aligned} \tag{6}$$

We can now substitute that cosine when computing the projection of $\overrightarrow{D_{Ip}}$ over the longitudinal axis of the red capsule ($\overrightarrow{D_{Pr}}$ in Figure 4):

$$\begin{aligned} |\overrightarrow{D_{Pr}}| &= cos(\beta) * |\overrightarrow{D_{Ip}}| \\ |\overrightarrow{D_{Pr}}| &= \frac{\overrightarrow{D_{Ip}} \cdot \overrightarrow{D_{CTc}}}{|\overrightarrow{D_{CTc}}| * |\overrightarrow{D_{Ip}}|} * |\overrightarrow{D_{Ip}}| \\ |\overrightarrow{D_{Pr}}| &= \frac{\overrightarrow{D_{Ip}} \cdot \overrightarrow{D_{CTc}}}{|\overrightarrow{D_{CTc}}|} \end{aligned} \tag{7}$$

Having that module, we only have to subtract $|\overrightarrow{D_{CTc}}|$ in order to obtain the desired distance:

$$\begin{aligned} ND(Ip, Sp, CTc) &= \frac{\overrightarrow{D_{Ip}} \cdot \overrightarrow{D_{CTc}}}{|\overrightarrow{D_{CTc}}|} - |\overrightarrow{D_{CTc}}| \\ ND(Ip, Sp, CTc) &= \frac{\overrightarrow{Ip - Sp} \cdot \overrightarrow{CTc - Sp}}{|\overrightarrow{CTc - Sp}|} - |\overrightarrow{CTc - Sp}| \end{aligned} \tag{8}$$

Computing the distance like this, with this final subtraction, allows to obtain a positive distance when impact point is outside the hand mesh, and a negative one if it is inside.

We compute the nearest distance per each capsule trigger attached to a finger phalanx. As stated before, if the distance is negative, this indicates a finger penetration issue on the object surface. Otherwise, if distance is positive, it means that finger stopped above the object surface. The ideal case is when a zero distance is obtained, that is, the finger is perfectly situated on the object surface.

The total error for the hand is represented by the following equation:

$$HandError = \sum_{i=1}^{N_{Fingers}} \frac{\sum_{j=1}^{N_{CTF}} |ND(Ip_{ij}, Sp_{ij}, CTc_{ij})|}{N_{CapsuleTriggersPerFinger}} \tag{9}$$

4.3 Dataset

To benchmark our grasping system we used a set of objects that are frequently used in daily life, such as, food items (e.g. cracker box, cans, box of sugar, fruits, etc.), tool items (e.g. power drill, hammer, screwdrivers, etc.), kitchen items (e.g. eating utensils) and also spherical shaped objects (e.g. tennis ball, racquetball, golf ball, etc.). Yale-CMU-Berkeley (YCB) Object and Model set [1] provides us these real-life 3D textured models scanned with outstanding accuracy and detail. Available objects have a wide variety of shapes, textures and sizes as we can see in Figure 5. The advantage of using real life objects is that the user already has a previous experience manipulating similar objects so he will try to grab and interact with the objects in the same way.

4.4 Participants

For the performance analysis, we recruited ten participants (8M/2F) from the local campus. Four of them have experience with VR applications. The rest are inexperienced virtual reality users. Participants will take part on both qualitative and quantitative evaluation. The performance analysis procedure will be described in the following subsection, indicating the concrete tasks to be performed by each participant.

4.5 Procedure

The system performance analysis begins with the quantitative evaluation. In this first phase, the user will be embodied in a controlled scenario[2] where 30 different objects will be spawned in a delimited area, with random orientation, and in the same order as represented in Figure 5. The user will try to grasp the object as he would do in real life and as quickly as possible. For each grasping, the system will compute the error metric and will also store the time spent by the user in grasping the object. The purpose of this first phase is to visually analyze grasping quality which is directly related to user expertise in VR environments and concretely with our grasping system. An experienced user would know system limits both when interacting with complex geometries or with large objects that would make it difficult to perform the grasp action quickly and naturally.

For the qualitative evaluation, the same user will be embodied in a photorealistic scenario changing mannequin hands by human hand model with realistic textures. After interacting freely in the photorealistic virtual environment[3], the user will have to answer the evaluation questionnaire defined in Table 1. The main purpose is the evaluation of interaction realism, finger and hand movement naturalness and motor control, among other qualitative aspects regarding user experience in VR environments.

5 Results and Discussion

In this section we will discuss and analyze the results obtained from the performance analysis process. On the one hand, we will draw conclusions from the average error obtained in grasping each object by each participant group, and also from the overall error per object taking into account all the participants (see Figure 7). On the other hand, we obtained the average elapsed time needed in grasping each object for each participant group, and also the average elapsed time needed for each object taking into account all the participants (see Figure 6). This will allow us to draw conclusions about the most difficult objects to manipulate in terms of accuracy and elapsed time for grasping. Moreover, we can compare system performance used by inexperienced users in comparison with experienced ones.

5.1 Qualitative evaluation

Qualitative evaluation for each participant was calculated using the Equation 3 obtaining a score for each qualitative aspect. In Table 2 we represent for each group of participants: the average score for each evaluation aspect and the total embodiment score computed using the Equation 4. Regarding the represented results in Table 2, the evaluation of experienced users has been more disadvantageous as they have a more elaborated criterion given their previous experience with virtual reality applications. Finger movement realism (aspect 2) was evaluated similarly by both groups. This is because the hand closing and opening gestures are guided by the same animation in both cases. Finally, the reported results referring to the interaction realism have been the lowest in both cases. This is mostly because users cannot control their individual fingers movement, since general hand gesture is controlled by a unique trigger button of the controller. However, overall obtained embodiment score is 2.08 out of 3.0.

	Score	
Evaluation Aspects	Experienced users	Inexperienced users
(1) Motor Control	1.85	2.34
(2) Finger Movement Realism	2.33	2.51
(3) Interaction Realism	1.84	1.95
Embodiment score	1.97	2.19

TABLE 2: Average score for each qualitative aspect of the evaluation and group of participants. Maximum result would be three.

5.2 Quantitative evaluation

As expected, inexperienced users have taken longer to grasp almost all the object set due to the lack of practice and expertise with the system. This is clearly represented in Figure 6 where experienced users only have taken longer in grasping some tools such as, the flat screwdriver (Figure 5z) and hammer (Figure 5aa). Inexperienced users take an average of 0.36 seconds longer to grab the objects. In practice and regarding interaction, this is not a factor that is going to make a crucial difference. Analyzing Figure 6, the tuna fish can (Figure 5f), potted meat can (Figure 5h), spatula (Figure 5u), toy airplane (Figure 5ad) and bleach cleaner (Figure 5q) are the most time consuming when grasped by the users. This is mainly because of their sizes and complex geometries. Since objects are spawned with a random orientation, this fact can affect grasping times. Even so, we can conclude that the largest objects are those that the user takes the longest to grasp.

Regarding Figure 7 we can observe that the errors obtained by both groups of participants are quite similar. Most significant differences were observed in the case of power drill (Figure 5v) and the spatula. The power drill has a complex geometry and its size also hinders its grasp as the same as spatula and toy airplane.

Analyzing the overall error in the Figure 7, we conclude that the largest objects, such as the toy airplane, power drill, and bleach cleaner are those reporting most error rate. In addition, we observe how overall error decreases from the first objects to the last ones. This is mainly because, user skills and expertise with the grasping system are improving progressively. Moreover, results refer to a steep learning curve.

2. https://youtu.be/4sPhLbHpywM
3. https://youtu.be/65gdFdwsTVg

(a) Master Chef can (b) Cracker box (c) Sugar box (d) Tomato soup can (e) Mustard bottle (f) Tuna fish can

(g) Gelatin box (h) Potted meat can (i) Banana (j) Strawberry (k) Apple (l) Lemon

(m) Peach (n) Pear (o) Orange (p) Plum (q) Bleach cleaner (r) Fork

(s) Spoon (t) Knife (u) Spatula (v) Power drill (w) Scissors (x) Large marker

(y) Adjustable wrench (z) Flat screwdriver (aa) Hammer (ab) Baseball (ac) Tennis ball (ad) Toy airplane

Fig. 5: Grasping performed on objects from the YCB dataset.

6 Applications

Our grasping system can be applied to several existing problems in different areas of interest, such as: robotics [26], rehabilitation [27] and interaction using augmented reality [28].

In robotics, different works have been explored to implement robust grasp approaches that allow robots to interact with the environment. These contributions are organized in mainly four different blocks [29]: methods that rely on known objects and previously estimated grasp points [30], grasping methods for familiar objects [31], methods for unknown objects based on the analysis of object geometry [32] and automatic learning approaches [33]. Our approach is more closely related to this last block, where its use would potentially be a relevant contribution. As a direct application, our system enables human-robot knowledge transfer where robots try to imitate human behaviour in performing grasping.

Our grasping system is also useful for rehabilitation of patients with hand motor difficulties, and it could even be done remotely assisted by an expert [34], or through an automatic system [35]. Several works have demonstrated the viability of patient rehabilitation in virtual environments [27], helping them to improve the mobility of their hands in daily tasks [36]. Our novel error metric in combination with other automatic learning methods, can be used to guide patients during rehabilitation with feedback information and instructions. This will make rehabilitation a more attractive process, by quantifying the patient progress and visualizing its improvements over the duration of rehabilitation.

Finally, our grasping system integrated in UnrealROX [2] enables many other computer vision and artificial intelligence applications by providing synthetic ground truth data, such as depth and normal maps, object masks, trajectories, stereo pairs, etc. of the virtual human hands interacting with real objects from the YCB dataset (Figure 8).

7 Limitations and Future Works

Our proposal has several major limitations:

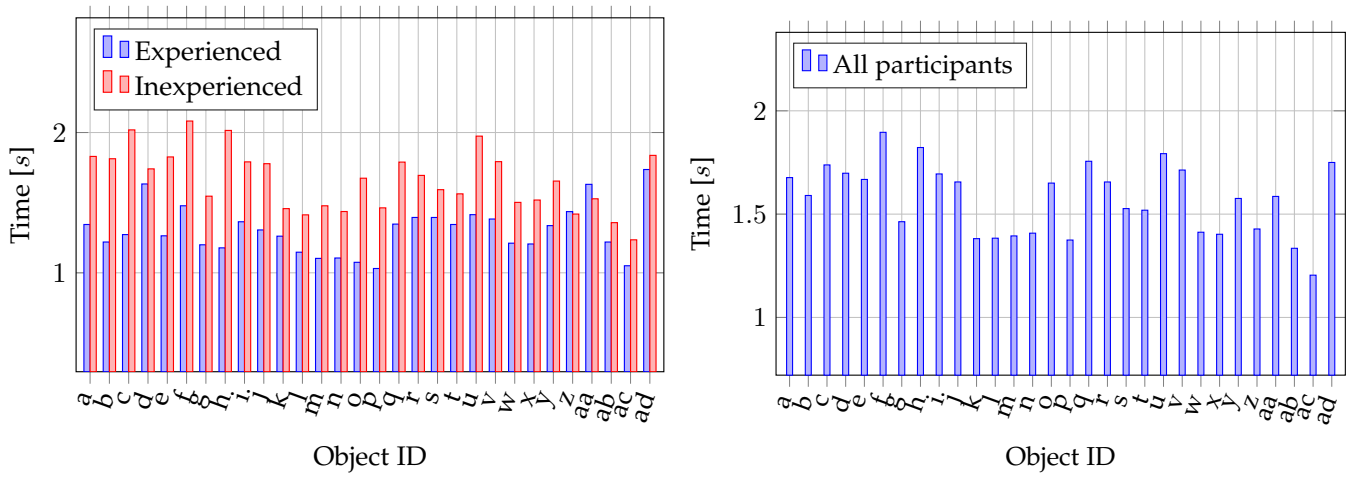


Fig. 6: At left, average elapsed time obtained by each participants group on grasping each object. At right, average elapsed time obtained by all the participants on grasping each object.

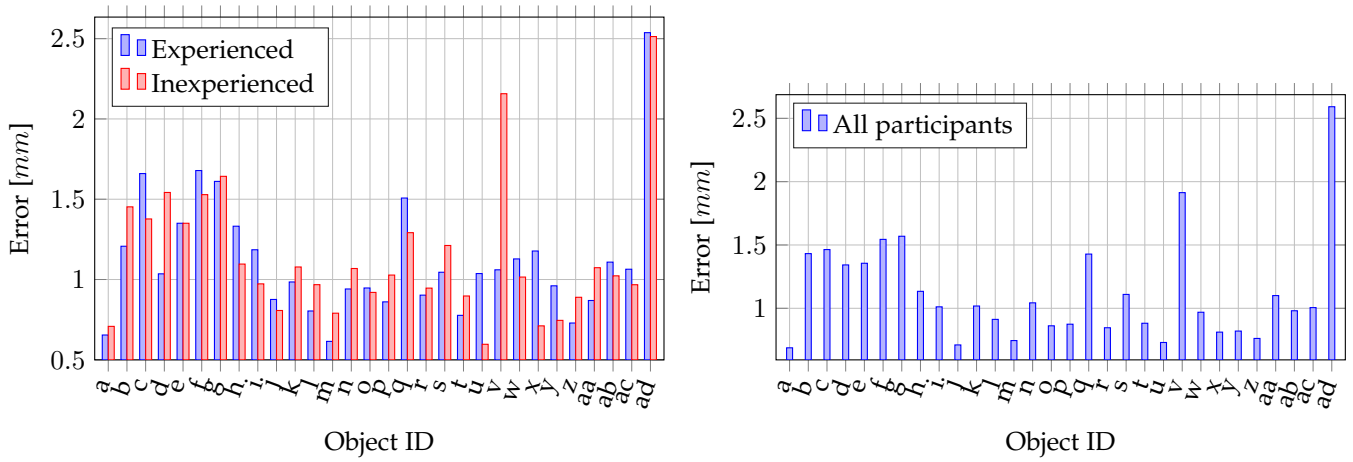
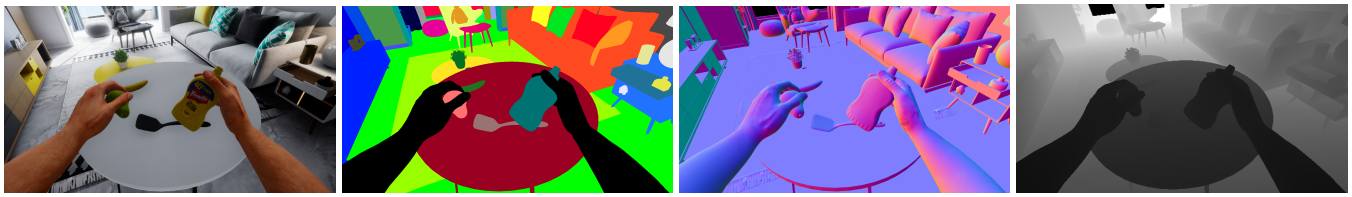


Fig. 7: At left, average error(mm) obtained by each participants group and for each object.At right, average error(mm) obtained by all the participants and for each object.



(a) RGB image. (b) Object mask. (c) Normal map. (d) Depth map.

Fig. 8: Different ground truth data extracted using UnrealRox alongside our grasping system.

- Hand movement is based on a single animation regardless object geometry. Depending on the object shape we could vary grasping gesture: spherical-grasping, cylindrical-grasping, finger-pinch, key-pinch, etc. However, our grasping gesture was experimentally the best when dealing with different shaped objects.
- The object can be grasped with only one hand. The user can interact with different objects using both hands at the same time. But not the same object with both hands.
- Sometimes it is difficult to deal with large objects due to the initial hand posture or because objects slide out from the hand palm due to physical collisions. Experienced users can better deal with this problem.

As future work, and in order to improve our grasping system, we could vary the hand grip gesture according to the object geometry we are manipulating. This is finding a correspondence between object geometry and a simple shape, e.g. a tennis ball is similar to a sphere thus proceeding with a spherical grasp movement. At the application level, there are several possibilities as we discussed in the previ-

ous section. However, we would like to emphasize the use of contact points obtained when grasping an object in virtual reality, to transfer that knowledge and human behavior to real robots.

8 Conclusion

This work proposes a flexible and realistic looking grasping system which enables smooth and real-time interaction in virtual reality environments with arbitrary shaped objects. This approach is unconstrained by the object geometry, it is fully controlled by the user and it is modular and easily implemented on different meshes or skeletal configurations. In order to validate our approach, an exhaustive evaluation process was carried out. Our system was evaluated qualitatively and quantitatively by two groups of participants: with previous experience in virtual reality environments (experienced users) and without expertise in VR (inexperienced). For the quantitative evaluation, a new error metric has been proposed to evaluate each grasp, quantifying hand-object overlapping. From the performance analysis results, we conclude that user overall experience was satisfactory and positive. Analyzing the quantitative evaluation, the error difference between experienced users and non experienced is subtle. Moreover, average errors are progressively smaller as more object are grasped. This clearly indicates a steep learning curve. In addition, the qualitative analysis points to a natural and realistic interaction. Users can freely manipulate previously defined dynamic objects in the photorealistic environment. Moreover, grasping contact points can be easily extracted, thus enabling numerous applications, especially in the field of robotics. Unreal Engine 4 project source code is available at GitHub alongside several video demonstrations. This approach can easily be implemented on different game engines.

Acknowledgments

This work has been funded by the Spanish Government TIN2016-76515-R grant for the COMBAHO project, supported with Feder funds. This work has also been supported by three Spanish national grants for PhD studies (FPU15/04516, FPU17/00166, and ACIF/2018/197), by the University of Alicante project GRE16-19, and by the Valencian Government project GV/2018/022. Experiments were made possible by a generous hardware donation from NVIDIA.

References

[1] B. Calli, A. Singh, A. Walsman, S. Srinivasa, P. Abbeel, and A. M. Dollar, "The ycb object and model set: Towards common benchmarks for manipulation research," in *Advanced Robotics (ICAR), 2015 International Conference on*. IEEE, 2015, pp. 510–517.

[2] P. Martinez-Gonzalez, S. Oprea, A. Garcia-Garcia, A. Jover-Alvarez, S. Orts-Escolano, and J. Garcia-Rodriguez, "Unrealrox: An extremely photorealistic virtual reality environment for robotics simulations and synthetic data generation," *arXiv preprint arXiv:1810.06936*, 2018.

[3] Y. Aydin and M. Nakajima, "Database guided computer animation of human grasping using forward and inverse kinematics," *Computers & Graphics*, vol. 23, no. 1, pp. 145–154, 1999.

[4] N. Wheatland, Y. Wang, H. Song, M. Neff, V. Zordan, and S. Jörg, "State of the art in hand and finger modeling and animation," in *Computer Graphics Forum*, vol. 34, no. 2. Wiley Online Library, 2015, pp. 735–760.

[5] F. Argelaguet and C. Andujar, "A survey of 3d object selection techniques for virtual environments," *Computers & Graphics*, vol. 37, no. 3, pp. 121–136, 2013.

[6] Y. Li, J. L. Fu, and N. S. Pollard, "Data-driven grasp synthesis using shape matching and task-based pruning," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 4, pp. 732–747, July 2007.

[7] C. Goldfeder and P. K. Allen, "Data-driven grasping," *Autonomous Robots*, vol. 31, no. 1, pp. 1–20, Jul 2011. [Online]. Available: https://doi.org/10.1007/s10514-011-9228-1

[8] P. Braido and X. Zhang, "Quantitative analysis of finger motion coordination in hand manipulative and gestic acts," *Human movement science*, vol. 22, no. 6, pp. 661–678, 2004.

[9] M. Ciocarlie, C. Goldfeder, and P. K. Allen, "Dimensionality reduction for hand-independent dexterous robotic grasping," 2007.

[10] S. Jörg and C. OSullivan, "Exploring the dimensionality of finger motion," in *Proceedings of the 9th Eurographics Ireland Workshop (EGIE 2009)*, 2009, pp. 1–11.

[11] N. S. Pollard and V. B. Zordan, "Physically based grasping control from example," in *Proceedings of the 2005 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, ser. SCA '05. New York, NY, USA: ACM, 2005, pp. 311–318. [Online]. Available: http://doi.acm.org/10.1145/1073368.1073413

[12] P. G. Kry and D. K. Pai, "Interaction capture and synthesis," in *ACM Transactions on Graphics (TOG)*, vol. 25, no. 3. ACM, 2006, pp. 872–880.

[13] Y. Bai and C. K. Liu, "Dexterous manipulation using both palm and fingers," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, May 2014, pp. 1560–1565.

[14] Y. Ye and C. K. Liu, "Synthesis of detailed hand manipulations using contact sampling," *ACM Transactions on Graphics (TOG)*, vol. 31, no. 4, p. 41, 2012.

[15] C. K. Liu, "Dextrous manipulation from a grasping pose," in *ACM SIGGRAPH 2009 Papers*, ser. SIGGRAPH '09. New York, NY, USA: ACM, 2009, pp. 59:1–59:6. [Online]. Available: http://doi.acm.org/10.1145/1576246.1531365

[16] J. Copenhaver, "Vr animation and locomotion systems in lone echo," pp. 17–27, 2017. [Online]. Available: https://www.gdcvault.com/play/1024446/It-s-All-in-the

[17] Oculus. Distance grab sample now available in oculus unity sample framework. [Online]. Available: https://developer.oculus.com/blog/distance-grab-sample-now-available-in-oculus-unity-sample-framework/

[18] ——. Oculus first contact. [Online]. Available: https://www.oculus.com/experiences/rift/1217155751659625/

[19] T. Looman. Vr template. [Online]. Available: https://wiki.unrealengine.com/VR_Template

[20] A. Christopoulos, M. Conrad, and M. Shukla, "Increasing student engagement through virtual interactions: How?" *Virtual Reality*, vol. 22, no. 4, pp. 353–369, Nov 2018. [Online]. Available: https://doi.org/10.1007/s10055-017-0330-3

[21] P. Koutsabasis and S. Vosinakis, "Kinesthetic interactions in museums: conveying cultural heritage by making use of ancient tools and (re-) constructing artworks," *Virtual Reality*, vol. 22, no. 2, pp. 103–118, Jun 2018. [Online]. Available: https://doi.org/10.1007/s10055-017-0325-0

[22] S. Vosinakis and P. Koutsabasis, "Evaluation of visual feedback techniques for virtual grasping with bare hands using leap motion and oculus rift," *Virtual Reality*, vol. 22, no. 1, pp. 47–62, Mar 2018. [Online]. Available: https://doi.org/10.1007/s10055-017-0313-4

[23] M. Gonzalez-Franco and T. C. Peck, "Avatar Embodiment. Towards a Standardized Questionnaire," *Frontiers in Robotics and AI*, vol. 5, p. 74, 2018. [Online]. Available: https://www.frontiersin.org/article/10.3389/frobt.2018.00074

[24] S. Poeschl and N. Doering, "The german vr simulation realism scale-psychometric construction for virtual reality applications with virtual humans." *Annual Review of Cybertherapy and Telemedicine*, vol. 11, pp. 33–37, 2013.

[25] D. E. Brackney and K. Priode, "Back to reality: the use of the presence questionnaire for measurement of fidelity in simulation," *J Nurs Meas*, vol. 25, no. 2, pp. 66–73, 2017.

[26] J. D. Bric, D. C. Lumbard, M. J. Frelich, and J. C. Gould, "Current state of virtual reality simulation in robotic surgery training: a review," *Surgical endoscopy*, vol. 30, no. 6, pp. 2169–2178, 2016.
[27] M. F. Levin, P. L. Weiss, and E. A. Keshner, "Emergence of virtual reality as a tool for upper limb rehabilitation: incorporation of motor control and motor learning principles," *Physical therapy*, vol. 95, no. 3, pp. 415–425, 2015.
[28] Z. Lv, A. Halawani, S. Feng, S. Ur Réhman, and H. Li, "Touch-less interactive augmented reality game on vision-based wearable device," *Personal and Ubiquitous Computing*, vol. 19, no. 3-4, pp. 551–567, 2015.
[29] J. Bohg, A. Morales, T. Asfour, and D. Kragic, "Data-driven grasp synthesisa survey," *IEEE Transactions on Robotics*, vol. 30, no. 2, pp. 289–309, 2014.
[30] Y. Lin and Y. Sun, "Robot grasp planning based on demonstrated grasp strategies," *The International Journal of Robotics Research*, vol. 34, no. 1, pp. 26–42, 2015.
[31] N. Vahrenkamp, L. Westkamp, N. Yamanobe, E. E. Aksoy, and T. Asfour, "Part-based grasp planning for familiar objects," in *Humanoid Robots (Humanoids), 2016 IEEE-RAS 16th International Conference on*. IEEE, 2016, pp. 919–925.
[32] B. S. Zapata-Impata, C. Mateo Agulló, P. Gil, and J. Pomares, "Using geometry to detect grasping points on 3d unknown point cloud," 2017.
[33] S. Levine, P. Pastor, A. Krizhevsky, J. Ibarz, and D. Quillen, "Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection," *The International Journal of Robotics Research*, vol. 37, no. 4-5, pp. 421–436, 2018.
[34] I. Escobar, C. Gálvez, G. Corrales, E. Pruna, M. Pilatasig, and J. Montaluisa, "Virtual system using haptic device for real-time tele-rehabilitation of upper limbs," in *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*. Springer, 2018, pp. 136–152.
[35] D. Avola, L. Cinque, G. L. Foresti, M. R. Marini, and D. Pannone, "Vrheab: a fully immersive motor rehabilitation system based on recurrent neural network," *Multimedia Tools and Applications*, pp. 1–28, 2018.
[36] A. L. Faria, A. Andrade, L. Soares, and S. B. i Badia, "Benefits of virtual reality based cognitive rehabilitation through simulated activities of daily living: a randomized controlled trial with stroke patients," *Journal of neuroengineering and rehabilitation*, vol. 13, no. 1, p. 96, 2016.

Sergiu Oprea is a PhD student in Computer Science at the University of Alicante. He received his Bachelor's Degree in Computer Engineering and his Master's Degree in Automation and Robotics from the same institution in June 2015 and June 2017, respectively. Currently, he is broadly interested in building the next generation of deep learning-based video prediction systems. His main research interests span topics mainly in computer vision, virtual/augmented Reality, and deep learning. He is also a member of European Networks like HiPEAC.

Pablo Martinez-Gonzalez is a PhD student in Computer Science at the University of Alicante specialized in online deep learning and object pose estimation. He received his Bachelor's Degree in Computer Engineering from the University of Alicante (Spain) in 2015 and his Master's Degree in Computer Graphics, Games and Virtual Reality from the University Rey Juan Carlos (Spain) in 2016. He is also interested in virtual reality and he is a member of European Networks such as HiPEAC.

Alberto Garcia-Garcia is a PhD student (Machine Learning and Computer Vision) at the University of Alicante. He received his Masters Degree (Automation and Robotics) and his Bachelors Degree (Computer Engineering) from the same institution in June 2016 and June 2015 respectively. His main research interests include deep learning (specially graph neural networks), 3D computer vision, and parallel computing on GPUs. He was an intern at Julich Supercomputing Center, intern at NVIDIA working jointly with the Camera/Solutions engineering team and the Mobile Visual Computing group from NVIDIA Research, and intern at Oculus Research (Facebook Reality Labs). He is also a member of European Networks such as HiPEAC and IV&L.

John Alejandro Castro-Vargas is a PhD student in Computer Science at the University of Alicante. He received his Bachelor's Degree in Computer Engineering and his Master's Degree in Automation and Robotics from the same institution in June 2016 and June 2017, respectively. His main research interests include deep learning applied to action recognition, robotic grasping and virtual reality. He is also a member of European Networks such as HiPEAC.

Sergio Orts-Escolano received a BSc, MSc and PhD in Computer Science from the University of Alicante (Spain) in 2008, 2010 and 2014 respectively. He is currently an assistant professor in the Department of Computer Science and Artificial Intelligence at the University of Alicante. Previously he was a researcher at Microsoft Research where he was one of the leading members of the Holoportation project (virtual 3D teleportation in real-time). His research interests include computer vision, 3D sensing, real-time computing, GPU computing, and deep learning. He has authored +50 publications in journals and top conferences like CVPR, SIGGRAPH, 3DV, BMVC, Neurocomputing, Neural Networks, Applied Soft Computing, etcetera. He is also member of European Networks like HiPEAC and Eucog.

Jose Garcia-Rodriguez received his Ph.D. degree, with specialization in Computer Vision and Neural Networks, from the University of Alicante (Spain). He is currently Associate Professor at the Department of Computer Technology of the University of Alicante. His research areas of interest include: computer vision, computational intelligence, machine learning, pattern recognition, robotics, man-machine interfaces, ambient intelligence, computational chemistry, and parallel and multicore architectures. He has authored +100 publications in journals and top conferences and revised papers for several journals like Journal of Machine Learning Research, Computational intelligence, Neurocomputing, Neural Networks, Applied Softcomputing, Image Vision and Computing, Journal of Computer Mathematics, IET on Image Processing, SPIE Optical Engineering and many others, chairing sessions in the last decade for WCCI/IJCNN and participating in program committees of several conferences including IJCNN, ICRA, ICANN, IWANN, IWINAC KES, ICDP and many others. He is also member of European Networks of Excellence and COST actions like Eucog, HIPEAC, AAPELE or I&VL and director or the GPU Research Center at University of Alicante and Phd program in Computer Science.