Fine Virtual Manipulation with Hands of Different Sizes

Suzanne Sorli, Dan Casas, Mickeal Verschoor, Ana Tajadura-Jiménez, Miguel A. Otaduy

Abstract—Natural interaction with virtual objects relies on two major technology components: hand tracking and hand-object physics simulation. There are functional solutions for these two components, but their hand representations may differ in size and skeletal morphology, hence making the connection non-trivial. In this paper, we introduce a pose retargeting strategy to connect the tracked and simulated hand representations, and we have formulated and solved this hand retargeting as an optimization problem. We have also carried out a user study that demonstrates the effectiveness of our approach to enable fine manipulations that are slow and awkward with naïve approaches.

Index Terms—Hand simulation, hand tracking, pose retargeting.

1 Introduction

To date, VR has reached a high degree of visual realism, allowing the creation of truly immersive virtual experiences [14, 15, 27]. When virtual objects appear real, the next natural step is to reach out and start interacting with them [6]. But this apparently simple action entails additional tasks in VR: hand tracking and hand-object simulation, which are typically solved independently. For hand tracking, the common solution is to use computer vision methods, which output the skeletal morphology and configuration of a hand that best matches the user's actual hand [20]. For hand-object simulation, the most general approach is to find the configuration of a simulated hand that takes the tracked hand as goal, but is subject to a model of hand biomechanics and the laws of contact mechanics [34]. Some modern commercial hand-tracking solutions, such as Oculus Quest, provide some limited hand interactions by building an ad-hoc physics-based model on top of the tracked hand morphology. In our work, we address challenges arising in the connection of hand tracking and hand-object simulation. As a result, we aim for VR animations of fine object manipulation, commanded by interactive hand tracking of the user's hands.

When connecting hand tracking and hand-object simulation, we find that the hand models used in these two tasks may differ in size and skeletal morphology. These differences may be due to at least two major reasons: First, it is non-trivial to produce a simulation model that fits exactly the size and morphology of the user's hand. Even though embodying the user in an avatar with different hand size is perfectly viable from a perceptual point of view [1], it is not free of technical difficulties. Second, to leverage existing work in hand tracking and hand simulation, it is convenient to integrate off-the-shelf solutions, but it is unlikely that these solutions use hand representations with the same skeletal morphology. For instance, there is no consensus on the placement of joints across different hand representations, particularly at the palm.

Due to these differences in hand size and skeletal morphology, the hand pose computed by hand tracking cannot be directly input to hand-object simulation. If the pose is applied naïvely, it results in inaccurate finger configurations, which complicate dexterous manipulation of virtual objects. Thanks to visual feedback of the simulated hand, the user may correct the real hand pose and try to work around the mismatch. We have found that this is sufficient for gross manipulation of virtual objects. However, some finger configurations are impossible to reach when the pose of the tracked hand is applied naïvely to the simulated hand, which altogether prevents dexterous fine manipulation of virtual objects.

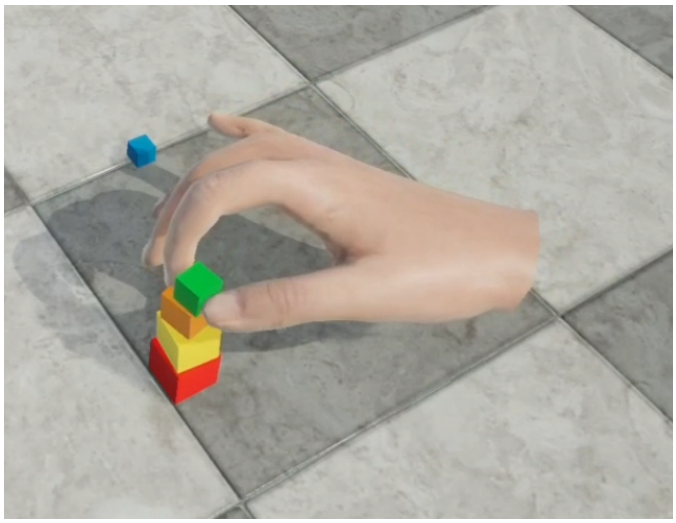

Fig. 1. Natural physics-based interaction with small objects. This VR scene was implemented by connecting off-the-shelf hand tracking and hand simulation solutions, which use hand models with different skeletal morphology.

In this paper, we introduce a pose retargeting strategy to connect the tracked hand and the simulated hand. Our approach works with any type of tracking or simulation method, as it stands at the interface between both tasks. We use an intermediate hand representation that shares the size and morphology of the simulated hand, but which tries to match the configuration of the tracked hand. The retargeting strategy formulates an optimization of the pose of this intermediate hand, based on features that represent the pose of the tracked hand. We have used finger tip positions as features, as they represent key information for fine manipulation. We describe the formulation and solution to the optimization problem in Section 3, together with a brief summary of the hand-object simulation using the CLAP library [34].

We have evaluated the practical impact of the hand mismatch on the manipulation of virtual objects, and we have compared our pose retargeting strategy vs. naïve copy of the hand pose. To this end, we have carried out a user study, discussed in Section 4, comparing task performance of virtual object manipulation. We have confirmed that the mismatch of the hand representation is not critical for gross manipulation (i.e., large objects), but it is critical for fine manipulation (i.e., small objects). With our pose retargeting approach, the performance of pick and drop actions for small objects is significantly faster than the performance of naïve strategies, and users also report increased precision, naturalness and ease of manipulation.

- *Suzanne Sorli, Dan Casas, Mickeal Verschoor and Miguel Otaduy are with the Computer Science Dept. at Universidad Rey Juan Carlos, Madrid, Spain. E-mail: {suzanne.sorli,dan.casas,mickeal.verschoor,miguel.otaduy}@urjc.es.*
- *Ana Tajadura-Jiménez is with the Computer Science Dept. at Universidad Carlos III de Madrid, Spain. E-mail: atajadur@inf.uc3m.es.*

2 Related Work

The computation of skeletal configurations of hand models is at the core of both hand tracking and hand simulation. These two lines of research differ in the input data and the formulation of the computational problem, but both solve the pose of the hand (i.e., joint or bone transformations).

Optical hand tracking takes as input images of hands or key feature points, and computes the skeletal configuration of the hand that best reproduces the input data. Modern methods can be classified into two large sets. Discriminative methods work by directly regressing the hand configuration based on the input data, and they require a training step [2, 5, 10, 16, 22, 33, 35]. Generative methods, on the other hand, work by finding the hand configuration that minimizes an objective function, and require a hand model but no training [19,28,32]. Our hand pose retargeting method shares the general methodology of generative hand tracking methods. Also related to ours are the recent tracking methods that are able to estimate the hand pose while manipulating rigid objects [26, 29]. However, physically-correct interactions cannot be enforced since forces are not modeled.

Physics-based hand simulation aims to compute a hand configuration that satisfies force equilibrium. The competing forces are dominated by contact and joint constraints, but may also include soft-tissue deformation. The different approaches consider articulated hand representations [3, 23], geometric flesh skinning [7], local skin deformation at fingers [13, 30], or full flesh deformation [8, 12]. The method of Verschoor et al. [34], which we use for our hand simulation, formulates the problem as an optimization. Our approach leverages the existing solutions for the tracking and simulation components, and formulates a simpler optimization problem whose goal is just to connect these two components in a simple way.

There is a broad line of research that studies the effect of the hand representation on the embodiment of the VR user [1]. Most works try to understand how different aspects of VR visualization and interaction affect embodiment, for example through analysis of the virtual hand illusion [17]. This line of research is orthogonal to our work. Its conclusions may indicate that embodiment is possible under notable differences in the simulated hand, and this calls for methods that bridge the tracked and simulated hand representations, such as our method.

While the focus of our work is hand simulation, the challenges and methods parallel those of skeletal body animation. Some authors have addressed the problem of motion retargeting across characters of very diverse morphology [11], or even within video-to-video [4].

3 Tracking-Based Hand Animation

As discussed in the introduction, we wish to drive a VR simulated hand model using as input interactive hand tracking data. However, the representations of the simulated hand and the tracked hand may differ in size and skeletal morphology. Furthermore, the simulated hand is constrained by contact with objects in the VR scene, while the tracked hand is not.

We use an intermediate hand representation to connect the user's tracked hand and the VR simulated hand. This intermediate hand shares some properties with the tracked hand (i.e., it is not constrained by other VR objects), and other properties with the simulated hand (i.e., its size and skeletal morphology). We characterize all three hand instances by their skeletal pose θ. Then, formally we denote the three following hand poses: θ^{t} for the user's tracked hand, θ^{i} for the intermediate hand representation, and θ^{s} for the VR simulated hand. Let us emphasize that, even though we use the same symbol θ to conceptually represent pose for all three hand instances, the joint angles of the tracked hand may have a different geometric interpretation from those of the intermediate and simulated hands, due to the differences in skeletal morphology. Recall that our method is general and works for any hand representation, hand tracker, and hand simulation method. In our implementation, we use Leap Motion as tracker, with its corresponding hand representation, and the MANO representation [25] for the simulated hand. Figure 2 shows schematically the interconnection of all three hand instances.

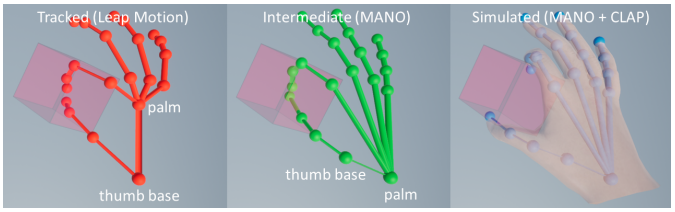

Fig. 2. Left: Tracked hand (in red), obtained using a Leap Motion tracker. Right: VR simulated hand (skeleton in blue, flesh semitransparent) implemented using CLAP [34] with MANO hand representation [25]. The skeletal morphology differs, in particular at the palm and thumb joints. Moreover, the simulated hand is constrained by contact with VR objects, while the tracked hand is not. Middle: We connect both hands using an intermediate representation (in green), which shares the morphology of the simulated hand but matches the pose of the tracked hand.

The intermediate hand representation serves as target configuration for the VR hand simulation. By matching the skeletal morphology of the simulated hand, it is easy to formulate input forces and torques for each bone in the simulated VR hand. These forces and torques are combined with contact forces and elastic deformation forces to produce the overall smooth simulation of the VR hand.

We start this section by describing a pose retargeting strategy to compute the intermediate hand θ^{i}. We motivate the formulation of the strategy to optimize fine manipulation tasks, and we describe an efficient solution algorithm. We conclude the section with a summary of the physics-based hand simulation method.

3.1 Hand Pose Retargeting

Given a pose of the tracked hand θ^{t}, we wish to compute a pose of the intermediate hand θ^{i}, such that it retains the most relevant characteristics, despite skeletal differences. We do this by defining a set of features $\mathbf{f}$, and solving an optimization problem. Prior to this, we apply a uniform scale to the tracked hand, such that it matches the overall size of the simulated hand. We do this by fitting a bounding box to an open palm pose.

The pose features should describe important characteristics of the pose, but with no assumptions about the skeletal morphology or size. In this work, we focus on the ability to manipulate small objects with high dexterity; therefore, we define the feature vector $\mathbf{f}$ by concatenating the positions of finger tips. In Section 4 we demonstrate that our pose retargeting approach is effective at producing dexterous manipulations of fine objects.

Based on this feature vector, we formulate the computation of the intermediate hand pose as the solution to the following constrained optimization problem:

$$\theta^{\mathrm{i}} = \arg\min_{\theta^{\mathrm{i}}} \frac{1}{2} \left(\mathbf{f}(\theta^{\mathrm{i}}) - \mathbf{f}(\theta^{\mathrm{t}})\right)^{\mathrm{T}} \mathbf{W} \left(\mathbf{f}(\theta^{\mathrm{i}}) - \mathbf{f}(\theta^{\mathrm{t}})\right) + R(\theta^{\mathrm{i}}), \quad (1)$$

$$\text{s.t. } \mathbf{c}(\theta^{\mathrm{i}}) \geq 0.$$

In a nutshell, the optimization finds the pose of the intermediate hand that produces features (i.e., finger tip positions) as close as possible to those of the tracked hand. $\mathbf{W}$ represents a (diagonal) weight matrix for the different features, which allows us to put more emphasis on the motion of the thumb and the index, for very accurate pinching. $R(\theta^{\mathrm{i}})$ is a regularization term; we use a small spatial and temporal regularization, to smooth interphalangeal rotations and avoid temporal discontinuities. $\mathbf{c}(\theta^{\mathrm{i}})$ represents constraints, to handle joint limits in the optimization.

We solve the optimization problem (1) iteratively using the Gauss-Newton method [21]. On each iteration, given a current estimate θ^{i}_0 of the pose of the intermediate hand, we linearize the feature vector as $\mathbf{f}(\theta^{\mathrm{i}}_0) + \frac{\partial \mathbf{f}}{\partial \theta^{\mathrm{i}}} \Delta\theta^{\mathrm{i}}$, and the active constraints as $\mathbf{c}(\theta^{\mathrm{i}}_0) + \frac{\partial \mathbf{c}}{\partial \theta^{\mathrm{i}}} \Delta\theta^{\mathrm{i}}$. Then, with Lagrange multipliers λ to enforce the active constraints, each iteration of Gauss-Newton amounts to solving the following linear

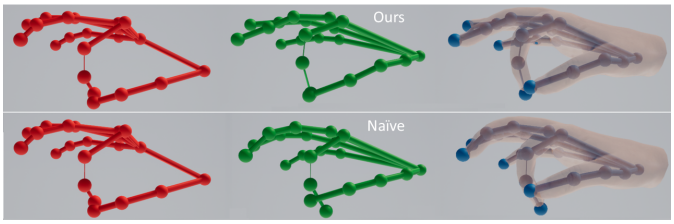

Fig. 3. With our pose retargeting (top), thumb-index pinch motions of the tracked hand are accurately reproduced on the simulated hand, despite skeletal differences. Naïve retargeting (bottom) does not reach comparable accuracy, which complicates fine manipulation.

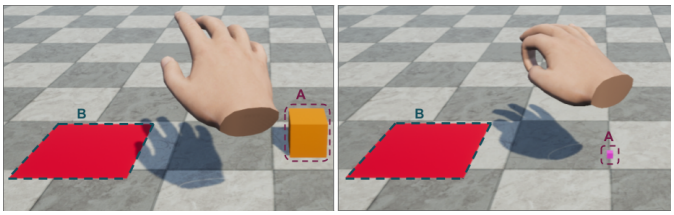

Fig. 4. We have studied a cube manipulation task, where users were asked to move a cube from A to B. In a user experiment, we have compared Our pose retargeting Strategy vs. a Naïve Strategy on two Manipulation scenarios: Gross Manipulation (i.e., large cube, left) and Fine Manipulation (i.e., small cube, right).

system:

$$\begin{pmatrix} \frac{\partial \mathbf{f}}{\partial \theta^{\mathrm{i}}}^{\mathrm{T}} \mathbf{W} \frac{\partial \mathbf{f}}{\partial \theta^{\mathrm{i}}} + \frac{\partial^2 R}{\partial \theta^{\mathrm{i}^2}} & \frac{\partial \mathbf{c}}{\partial \theta^{\mathrm{i}}}^{\mathrm{T}} \\ \frac{\partial \mathbf{c}}{\partial \theta^{\mathrm{i}}} & 0 \end{pmatrix} \begin{pmatrix} \Delta\theta^{\mathrm{i}} \\ \lambda \end{pmatrix} = \begin{pmatrix} \frac{\partial \mathbf{f}}{\partial \theta^{\mathrm{i}}}^{\mathrm{T}} \mathbf{W} \left(\mathbf{f}(\theta^{\mathrm{t}}) - \mathbf{f}(\theta_0^{\mathrm{i}})\right) - \frac{\partial R}{\partial \theta^{\mathrm{i}}}^{\mathrm{T}} \\ -\mathbf{c}(\theta_0^{\mathrm{i}}) \end{pmatrix}. \quad (2)$$

To solve the linear system, we compute a Cholesky factorization of the matrix $\frac{\partial \mathbf{f}}{\partial \theta^{\mathrm{i}}}^{\mathrm{T}} \mathbf{W} \frac{\partial \mathbf{f}}{\partial \theta^{\mathrm{i}}} + \frac{\partial^2 R}{\partial \theta^{\mathrm{i}^2}}$, then we use a Schur-complement approach to compute the Lagrange multipliers λ, and we conclude by computing the pose update $\Delta\theta^{\mathrm{i}}$.

Let us pay some attention to the computation of gradients $\frac{\partial \mathbf{f}}{\partial \theta^{\mathrm{i}}}$. Without loss of generality, finger tip positions $\mathbf{f}_j$ can be computed from the hand pose as a concatenation of relative rotations:

$$\mathbf{f}_j = \mathbf{x}_{\mathrm{wrist}} + \mathbf{R}_{\mathrm{palm}} \left(\mathbf{x}_{\mathrm{palm}} + \mathbf{R}_1 \left(\mathbf{x}_1 + \mathbf{R}_2 \left(\mathbf{x}_2 + \mathbf{R}_3 \mathbf{x}_3\right)\right)\right). \quad (3)$$

$\mathbf{x}_k$ and $\mathbf{R}_k$ denote, respectively, phalanx bone vectors and joint rotations. For gradient computations, we use a tangent-space representation of rotations [31]. In a nutshell, given the current joint rotation $\mathbf{R}_0$, a joint axis $\mathbf{k}$, and a differential rotation angle ϕ, the joint rotation can be expressed as $\mathbf{R} = (\mathbf{I} + \phi\,\mathrm{skew}(\mathbf{k}))\,\mathbf{R}_0$. The gradient with respect to the rotation angle is then simply $\frac{\partial \mathbf{R}}{\partial \phi} = \mathrm{skew}(\mathbf{k})\,\mathbf{R}_0$. We use this expression to compute the gradient of finger tip positions with respect to each component of the hand pose in (3), and thus we assemble the full gradient $\frac{\partial \mathbf{f}}{\partial \theta^{\mathrm{i}}}$ to be used in (2). In some joints, rotations have two degrees of freedom, and are expressed as the concatenation of two one-degree-of-freedom rotations.

Figure 3 depicts the accuracy of our pose retargeting strategy on a thumb-index pinch pose. We also compare the accuracy of a naïve retargeting strategy. For this, we align the intermediate and tracked hands using the base positions of the four fingers (which are very similar on the Leap Motion and MANO skeletons), we place the palm and thumb base joints based on the dimensions of the MANO skeleton, and then we copy the relative Leap Motion pose to the intermediate MANO hand. As shown in the figure, the naïve strategy fails to reproduce pinch poses correctly, which complicates fine manipulation.

The constrained Gauss-Newton solver requires on average 16 iterations per frame. This amounts to a total cost of 5 ms. per frame in our implementation on a commodity processor.

3.2 Hand Simulation

On every simulation frame, once the pose of the intermediate hand is computed following the retargeting approach described above, we use it to command the simulation of the VR hand. As mentioned in the introduction, we use the freely available CLAP simulation library to this end [34]. The decomposition of the hand animation into two subproblems, retargeting and simulation, simplifies the overall approach, and allows us to leverage off-the-shelf simulation libraries. As a result, we can create with no effort VR scenes with physics-based contact and natural hand interaction, as the one shown in Figure 1.

The hand simulation works as follows. Its degrees of freedom are: the positions and orientations of bones (gathered as the pose of the simulated hand, θ^{s}), the nodes of a tetrahedral decomposition of the skin, and the position and orientation of a grasped object (if it exists). These degrees of freedom are computed jointly on each simulation frame, by solving an optimization-like formulation of implicit integration of the deformation dynamics [9, 18]. The formulation of dynamics gathers multiple potential energy terms, which model the physical behavior of the hand, its interaction with the grasped object, and also the command provided by hand tracking. The energy terms are: soft-tissue deformation energy of the skin, including a nonlinear term for hyperelasticity; soft-constraint formulations of skeletal joints plus joint limits; skin-skeleton coupling; non-penetration and frictional contact constraints with the grasped object; and quadratic energy terms to penalize the deviation between the simulated pose θ^{s} and the intermediate pose θ^{i}.

In practice, to set up the hand simulation, we first generate the rest-shape geometry of the hand surface and the skeleton based on particular values of the shape parameters of the MANO model [25]. Then, we tetrahedralize the volume of the hand, connect internal nodes of the tetrahedral mesh to the bones, and pass this simulation model to the CLAP library [34]. The simulation model consists of 16 bones and 2,291 tetrahedra. The cost of the physics-based simulation is 51 ms. on average per frame. This adds some latency, but it did not seem to affect the quality of interaction.

4 User Experiment

To evaluate our hand pose retargeting strategy and compare it to naïve retargeting, we have designed a user experiment. In this experiment, users must execute a manipulation task, and the results confirm that our pose retargeting strategy enables more effective manipulation of virtual objects when fine dexterity is needed. In this section, we describe the experiment and discuss the results.

4.1 Methods

Participants. A total of 20 right-handed participants (age in years: range = 18-34, M = 26.2, SD = 4.04; 15 male and 5 female) took part in the user study. They received no compensation. In addition to age, we documented their hand size (range = 15.9-20.5 cm, M = 18.80, SD = 1.23) and prior VR experience (10 participants had experience, 10 had none), in order to test the influence of these variables. All participants confirmed correct vision with the HMD. The study was conducted in accordance with the 1964 Declaration of Helsinki and was granted ethical approval by the local ethics committee at Universidad Rey Juan Carlos. All participants provided informed written consent beforehand.

Materials and experimental design. We studied a manipulation task where users were asked to pick a cube with their thumb and index finger from position A, move it to position B, and drop it there, as shown in Figure 4. The interaction between the user's hand and the cube, as well as the interaction of the cube with the floor, were simulated with full physics-based contact, as described in Section 3.2. However, we disabled contact between the simulated hand and the floor. In the absence of haptic feedback, we found that the response of floor-hand contact could be unintuitive and distort the experiment. The scenarios

were developed in Unreal Engine, and were displayed on an Oculus Rift HMD, with head tracking, to optimize vision-motion correlation and make the manipulation task very natural. The hand of the user was tracked using a Leap Motion device, mounted frontally on the HMD for optimal tracking accuracy of grasping and pinching poses. The HMD was sanitized after each use, and soft components were covered with disposable hygienic pads.

The simulated VR hand was the same throughout the experiment. It was a large hand, 21.4 cm long, generated by adjusting the value of the main component of the statistical MANO model [25]. Therefore, the skeletal morphology of both the simulated and intermediate hands corresponded to the MANO model. The skeletal morphology of the tracked hand corresponded instead to the morphology of the Leap Motion model. Moreover, the size of the tracked hand was adapted to each user, thanks to the built-in functionality of the Leap Motion.

Two different strategies were compared in the study: the pose retargeting strategy described in Section 3.1, referred to as Ours, and a naïve retargeting strategy carried out by aligning the palms of the tracked and intermediate hands and then directly copying the joint angles of the tracked hand to the intermediate hand, referred to as Naïve.

In addition to the retargeting strategy, two different manipulation scenarios were studied: manipulation of a large cube 6 cm wide (i.e. Gross Manipulation) and a small cube 1 cm wide (i.e. Fine Manipulation), as shown in Figure 4.

Hypothesis. The initial hypothesis of the study is that the retargeting strategy may have an effect on the dexterity of manipulation; therefore, it may affect task performance on the Fine Manipulation scenario in which the small cube is manipulated, and to a lesser or no extent on the Gross Manipulation scenario in which the large cube is manipulated.

Experimental procedure. In each experimental trial, users were asked to execute the cube manipulation task. Each participant tested both types of Manipulation (i.e. Gross vs. Fine) under both Strategy conditions (i.e. Naïve vs. Ours), a total of five trials per Manipulation and Strategy. Having five repetitions of the condition allowed us to evaluate the effect of task learning. Participants completed five experimental blocks, each with four trials, one per experimental condition; on each block, the order of the four combinations of Manipulation and Strategy was randomized, to avoid potential bias due to task learning. Each experimental block lasted on average six minutes, and the full procedure lasted 30 minutes.

Measures and questionnaire. The time needed to complete the manipulation task in each experimental condition was measured. In addition, participants were asked to complete a questionnaire during the last experimental block, after each trial, i.e., once per condition. The questionnaire contained three items (5-point Likert-type), and was used to assess participants' subjective feelings of dexterity of manipulation in each experimental condition, in terms of Precision (defined to participants as "the movements of the virtual hand respond precisely to the movements made in reality", and ranging from "Not precise at all" to "Very precise"), Ease (defined to participants as "repetitions needed to achieve the objective", and ranging from "A lot of effort" to "Very little effort"), and Naturalness (defined to participants as "the way of grasping virtual objects corresponds to the way of grasping objects in the real world" and ranging from "Not at all natural" to "Very natural").

Data analysis. Data were statistically analyzed using R software. Time data was analyzed with repeated measures analyses of variance (ANOVA) with 2x2x5 within-subject factors Manipulation, Strategy and Repetition. In case of significant interactions between factors, these were followed by t-tests comparing all conditions against each other to understand if there were differences between them, with the p-value adjusted with the recommended Tukey method for comparing a family of estimates [36].

For questionnaire data, we conducted non-parametric Friedman tests to assess significant differences between the four conditions (i.e., Gross and Fine Manipulation under both Strategy conditions, i.e., Naïve vs. Ours). Significant results were followed by pairwise comparisons using Wilcoxon signed-rank tests comparing all four conditions against each other, with the p-value adjusted using the Bonferroni multiple testing correction method.

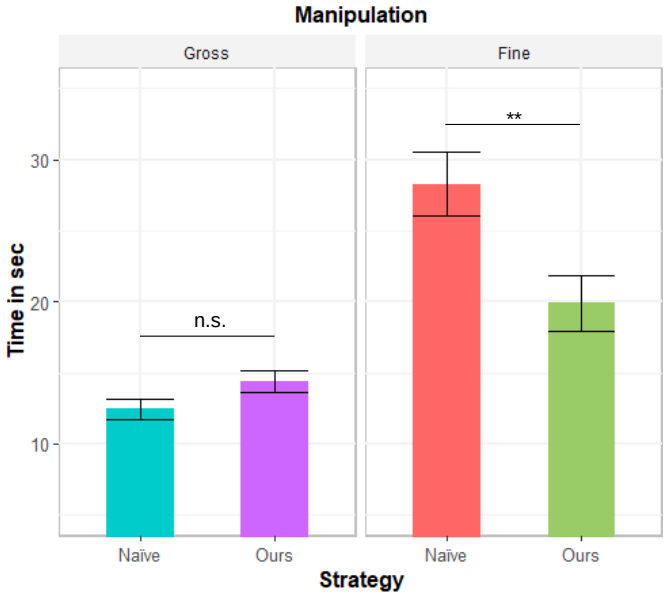

Fig. 5. Mean (±SE) time to complete the task for the four experimental conditions. Asterisks denote significant differences between means (** denotes $p<0.01$). n.s. denotes no significant differences between means. Note that only the relevant comparisons are displayed in the figure; results from all comparisons are described in the main text.

4.2 Results and Analysis

Time to complete the task. As shown in Figure 5, the Manipulation condition influenced the time to complete the task, but this influence was different depending on the Strategy. The ANOVA on the time data showed a significant difference between Manipulation conditions ($F(1,19)=15.21$, $p<0.001$), due to longer times needed to complete the task for the Fine than for the Gross Manipulation. Critically, there was a significant interaction between Manipulation and Strategy ($F(1,19)=8.78$, $p=0.008$). T-tests comparing the four conditions against each other showed the expected significant difference between the Gross and Fine Manipulation for the Naïve Strategy ($t(19)=4.881$, $p<0.001$), but this difference did not reach significance for Our Strategy ($p=0.34$). Importantly, t-tests also revealed that, while there were no significant differences between Strategies for the Gross Manipulation ($p=0.85$), the time to complete the task for the Fine Manipulation was significantly smaller for Our Strategy than for the Naïve Strategy condition ($t(19)=3.52$, $p=0.006$). This result indicates that the Fine Manipulation was easier to perform with Our Strategy. This is further evidenced by the results showing a significant difference between the Fine Manipulation with Naïve Strategy vs. Gross Manipulation with Our Strategy ($t(19)=4.36$, $p=0.001$), but not between the Gross Manipulation with Naïve Strategy vs. Fine Manipulation with Our Strategy ($p=0.11$).

Regarding the effect of Repetition, there was a significant main effect ($F(4,76)=10.01$, $p<0.001$), showing an effect of learning with more repetitions. Further, a significant interaction of Repetition with the factor Manipulation ($F(4,76)=4.71$, $p=0.002$) was found; as it can be seen in Figure 6-left, the effect of learning with more repetitions was larger for the Fine than for the Gross Manipulation, because of the task being easier for the Gross Manipulation, which was expected. Importantly, there was also a significant interaction of Repetition with the factor Strategy ($F(4,76)=3.69$, $p=0.008$), showing that the effect of learning with more repetitions was larger for the Naïve Strategy than for Our Strategy. As it can be seen in Figure 6-right, for the very first trial it took less time to perform the task with Our Strategy than with the Naïve strategy ($t(19)=3.82$, $p=0.01$). This suggests that Our Strategy makes the task easier and more natural than the Naïve Strategy. There was not a significant triple interaction between the

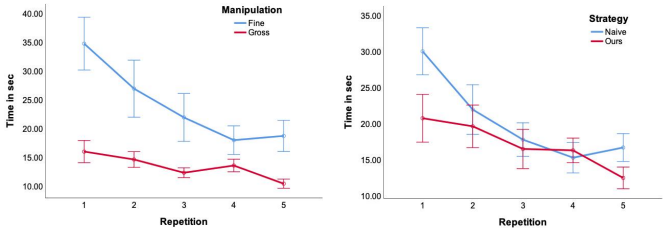

Fig. 6. Mean (±SE) time to complete the task across repetitions for the two Manipulation conditions (left) and for the two Strategy conditions (right).

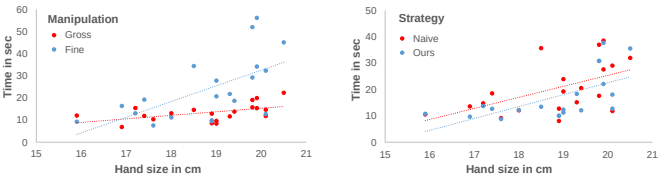

Fig. 7. Mean (±SE) time to complete the task for the two Manipulation conditions (left) and for the two Strategy conditions (right) according to participant's hand size.

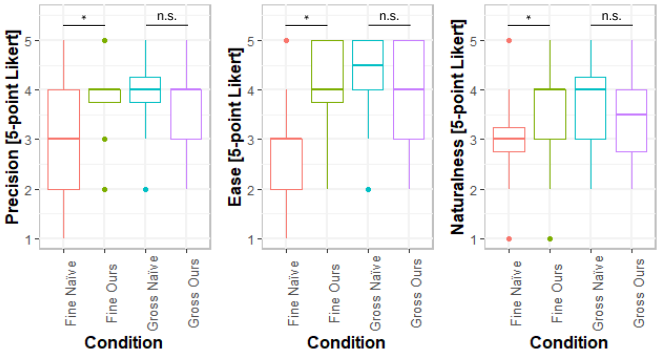

Fig. 8. Median (±Range) self-reported scores for Precision, Ease and Naturalness. Asterisks denote significant differences between means (* denotes $p < 0.05$); n.s. denotes no significant differences between means. Note that only the relevant comparisons are displayed in the figure; results from all comparisons are described in the main text.

factors Repetition, Manipulation and Strategy, accounting for different effects of Repetition for the two types of Manipulation with the two different Strategies.

To test the potential influence of the individual factors age, hand size and prior VR experience on the observed effects, we ran additional ANOVAs with factors Manipulation and Strategy in which we added age and hand size as covariates, and prior VR experience as a between-subjects factor. We observed a significant interaction of the factor Manipulation only for the factor hand size ($F(1,18)=8.38$, $p=0.01$). As shown in Figure 7, while overall it took less time for participants with smaller hand sizes to complete the task, this facilitation effect for smaller hand sizes was more evident for the Fine Manipulation condition. Importantly, the factors hand size, VR experience and age did not interact significantly with the factor Strategy (all ps$>$0.12), which indicates that the results related to Strategy were not significantly affected by these factors.

Self-report measures. Figure 8 shows the participants' subjective feelings of dexterity of manipulation, in terms of Precision, Ease and Naturalness, for each of the four experimental conditions. The precision score was statistically significantly different across conditions ($X^2(3)=14.43$, $p=0.002$). This was also the case for the effort score ($X^2(3)=15.36$, $p=0.001$) and the naturalness score ($X^2(3)=12.34$, $p=0.006$).

In terms of precision, pairwise Wilcoxon signed rank tests comparing the four conditions against each other showed that participants felt more precise in the Fine Manipulation with Our Strategy than with the Naïve Strategy ($p=0.017$), as shown in Figure 8-left. Further, for the Naïve Strategy, the Fine Manipulation felt significantly less precise than the Gross Manipulation ($p=0.021$), while this was not the case for Our Strategy, for which there were not significant differences between Manipulation conditions. Other comparisons between conditions were not significant either.

In terms of effort, pairwise Wilcoxon signed rank tests comparing the four conditions against each other showed that participants felt they had applied significantly less effort in the Fine Manipulation with Our Strategy than with the Naïve Strategy ($p=0.013$), as shown in Figure 8-middle. The Fine Manipulation with the Naïve Strategy required also more effort than the Gross Manipulation both with the Naïve Strategy ($p=0.008$) and with Our Strategy ($p=0.016$). Other comparisons between conditions were not significant.

In terms of naturalness, pairwise Wilcoxon signed rank tests comparing the four conditions against each other showed that participants reported higher naturalness of the Fine Manipulation with Our Strategy than with the Naïve Strategy ($p=0.047$), as shown in Figure 8-right. Further, for the Naïve Strategy, the Fine Manipulation felt significantly less natural than the Gross Manipulation ($p=0.019$), while this was not the case for Our Strategy, for which there were not significant differences between Manipulation conditions. Other comparisons between conditions were not significant either.

4.3 Discussion

The analysis of task performance and self-reporting of the user experiment suggests benefits of the proposed pose retargeting strategy. Moreover, these benefits appear independent of hand size, VR experience or age. First and foremost, Our Strategy exhibits significantly better performance than the Naïve Strategy for Fine Manipulation. In addition, the Naïve Strategy performs significantly worse on Fine Manipulation vs. Gross Manipulation, while Our Strategy does not exhibit a significant performance difference on these two Manipulation conditions.

Hand size has an effect on performance for Fine Manipulation regardless of the retargeting Strategy, but Our Strategy performs better than the Naïve Strategy consistently across hand sizes.

The questionnaires suggest that Our pose retargeting Strategy feels significantly more precise, easier, and more natural for Fine Manipulation. For Gross Manipulation, Our Strategy scores slightly lower than the Naïve Strategy, but the difference is not significant. This is likely due to the inherent easiness of the Gross Manipulation scenario, which is confirmed when analyzing task performance across repetitions: Fine Manipulation benefits from learning more significantly than Gross Manipulation. Similarly, the analysis of task performance across repetitions indicates that the performance gain of Our Strategy is even larger initially, which again suggests that it is more natural, i.e., it requires less training.

5 Limitations and Future Work

In this paper, we have proposed a method to retarget hand poses between hands with different size and skeletal morphology. The method serves for connecting off-the-shelf solutions for hand tracking and physics-based hand simulation, avoiding the need to share a common hand representation. The results of the user study indicate that our method is effective for fine manipulation, achieving performance and naturalness comparable to gross manipulation. From an applied point-of-view, our hand retargeting approach could accelerate the development of VR training applications requiring high dexterity and fine manipulation.

The key technical insight of the method is to formulate pose retargeting as the optimization of finger tip positions. This approach is motivated by maximizing the accuracy of pinch poses, which are key for fine manipulation. One interesting avenue of future work would

be to explore more diverse feature vectors and/or objective functions, including other points in the hand, pose likelihood, etc. Similarly, the user study could be extended by covering more diverse interaction tasks.

Our hand retargeting method assumes that the shape of the simulated VR hand is given. This is the case when the VR application uses a hand of a fixed size, but one interesting extension would be to allow the simulation of personalized hands, which would require changing the shape of the simulated hand. Our current approach approximates this step by estimating a uniform scale, which could be extended to the estimation of, e.g., statistical shape parameters [25].

Currently, the retargeting method is applicable only to hands with similar skeletal topology, e.g, with five fingers. However, the approach could be extended to connect hands with very diverse skeletons, e.g., with a different number of fingers. The technical challenge is to define relevant feature metrics for such diverse hands.

Acknowledgments

We would like to thank the reviewers for their feedback, Jiayi Wang for the hand texture [24], Jessica Illera for the help with the study, and other members of the MSLab at URJC for their support. This work was funded in part by the European Research Council (ERC Consolidator Grant no.772738 TouchDesign) and the Spanish Ministry of Science (RTI2018-098694-B-I00 VizLearning; PID2019-105579RB-I00/AEI/10.13039/501100011033).

References

[1] F. Argelaguet, L. Hoyet, M. Trico, and A. Lecuyer. The role of interaction in virtual embodiment: Effects of the virtual hand representation. In *2016 IEEE Virtual Reality (VR)*, pp. 3–10, 2016.

[2] S. Baek, K. I. Kim, and T.-K. Kim. Pushing the envelope for rgb-based dense 3d hand pose estimation via neural rendering. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[3] C. W. Borst and A. P. Indugula. Realistic virtual grasping. In *Proc. of IEEE Virtual Reality Conference*, 2005.

[4] C. Chan, S. Ginosar, T. Zhou, and A. A. Efros. Everybody dance now. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.

[5] Y. Chen, Z. Tu, L. Ge, D. Zhang, R. Chen, and J. Yuan. So-handnet: Self-organizing network for 3d hand pose estimation with semi-supervised learning. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.

[6] M. Chessa, G. Maiello, L. K. Klein, V. C. Paulun, and F. Solari. Grasping objects in immersive virtual reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 1749–1754, 2019.

[7] C. Duriez, H. Courtecuisse, J.-P. d. Plata Alcalde, and P.-J. Bensoussan. Contact skinning. *Eurographics conference (short paper)*, 2008.

[8] C. Garre, F. Hernandez, A. Gracia, and M. A. Otaduy. Interactive simulation of a deformable hand for haptic rendering. In *Proc. of World Haptics Conference*, 2011.

[9] T. F. Gast, C. Schroeder, A. Stomakhin, C. Jiang, and J. M. Teran. Optimization integrator for large time steps. *IEEE Transactions on Visualization and Computer Graphics*, 21(10):1103–1115, Oct 2015. doi: 10.1109/TVCG.2015.2459687

[10] L. Ge, Y. Cai, J. Weng, and J. Yuan. Hand pointnet: 3d hand pose estimation using point sets. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[11] C. Hecker, B. Raabe, R. W. Enslow, J. DeWeese, J. Maynard, and K. van Prooijen. Real-time motion retargeting to highly varied user-created morphologies. *ACM Trans. Graph.*, 27(3):1–11, 2008.

[12] K. Hirota and K. Tagawa. Interaction with virtual object using deformable hand. In *IEEE Virtual Reality (VR)*, pp. 49–56. IEEE, 2016.

[13] J. Jacobs and B. Froehlich. A soft hand model for physically-based manipulation of virtual objects. In *2011 IEEE Virtual Reality Conference (VR)*, pp. 11–18, Mar. 2011.

[14] A. Kaplan, J. Cruit, M. Endsley, S. Beers, B. Sawyer, and P. Hancock. The effects of virtual reality, augmented reality, and mixed reality as training enhancement methods: A meta-analysis. *Human Factors*, 63(4):706–726, 2021.

[15] M. Krichenbauer, G. Yamamoto, T. Taketom, C. Sandor, and H. Kato. Augmented reality versus virtual reality for 3d object manipulation. *IEEE Transactions on Visualization and Computer Graphics*, 24(2):1038–1048, 2018.

[16] S. Li and D. Lee. Point-to-pose voting based hand pose estimation using residual permutation equivariant layer. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[17] L. Lin, A. Normoyle, A. Adkins, Y. Sun, A. Robb, Y. Ye, M. Di Luca, and S. Jörg. The effect of hand size and interaction modality on the virtual hand illusion. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 510–518, 2019.

[18] S. Martin, B. Thomaszewski, E. Grinspun, and M. Gross. Example-based elastic materials. *ACM Trans. Graph.*, 30(4), 2011.

[19] S. Melax, L. Keselman, and S. Orsten. Dynamics based 3d skeletal hand tracking. In *Proceedings of Graphics Interface 2013*, GI '13, p. 63–70. Canadian Information Processing Society, CAN, 2013.

[20] F. Mueller, M. Davis, F. Bernard, O. Sotnychenko, M. Verschoor, M. A. Otaduy, D. Casas, and C. Theobalt. Real-time Pose and Shape Reconstruction of Two Interacting Hands With a Single Depth Camera. *ACM Transactions on Graphics (TOG)*, 38(4), 2019.

[21] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, New York, NY, USA, second ed., 2006.

[22] M. Oberweger, P. Wohlhart, and V. Lepetit. Training a feedback loop for hand pose estimation. In *IEEE International Conference on Computer Vision (ICCV)*, pp. 3316–3324, 2015.

[23] R. Ott, F. Vexo, and D. Thalmann. Two-handed haptic manipulation for CAD and VR applications. *Computer Aided Design & Applications*, 7(1), 2010.

[24] N. Qian, J. Wang, F. Mueller, F. Bernard, V. Golyanik, and C. Theobalt. HTML: A Parametric Hand Texture Model for 3D Hand Reconstruction and Personalization. In *Proceedings of the European Conference on Computer Vision (ECCV)*. Springer, 2020.

[25] J. Romero, D. Tzionas, and M. J. Black. Embodied Hands: Modeling and Capturing Hands and Bodies Together. *ACM Transactions on Graphics, (Proc. SIGGRAPH Asia)*, 36(6), Nov. 2017.

[26] S. Sridhar, F. Mueller, M. Zollhöfer, D. Casas, A. Oulasvirta, and C. Theobalt. Real-time joint tracking of a hand manipulating an object from rgb-d input. In *European Conference on Computer Vision*, pp. 294–310. Springer, 2016.

[27] Q. Sun, A. Patney, L.-Y. Wei, O. Shapira, J. Lu, P. Asente, S. Zhu, M. Mcguire, D. Luebke, and A. Kaufman. Towards virtual reality infinite walking: Dynamic saccadic redirection. *ACM Trans. Graph.*, 37(4), 2018.

[28] A. Tagliasacchi, M. Schröder, A. Tkach, S. Bouaziz, M. Botsch, and M. Pauly. Robust articulated-icp for real-time hand tracking. *Computer Graphics Forum*, 34(5):101–114, 2015.

[29] O. Taheri, N. Ghorbani, M. J. Black, and D. Tzionas. Grab: A dataset of whole-body human grasping of objects. In *European Conference on Computer Vision*, pp. 581–600. Springer, 2020.

[30] A. Talvas, M. Marchal, C. Duriez, and M. A. Otaduy. Aggregate constraints for virtual manipulation with soft fingers. *IEEE Transactions on Visualization and Computer Graphics*, 21(4):452–461, 2015.

[31] C. J. Taylor and D. J. Kriegman. Minimization on the Lie Group SO(3) and related manifolds. Technical report, Yale University, 1994.

[32] A. Tkach, M. Pauly, and A. Tagliasacchi. Sphere-meshes for real-time hand modeling and tracking. *ACM Trans. Graph.*, 35(6), 2016.

[33] J. Tompson, M. Stein, Y. Lecun, and K. Perlin. Real-time continuous pose recovery of human hands using convolutional networks. *ACM Transactions on Graphics*, 33, August 2014.

[34] M. Verschoor, D. Lobo, and M. A. Otaduy. Soft-Hand Simulation for Smooth and Robust Natural Interaction. In *Proc. of the IEEE Virtual Reality Conference*, 2018.

[35] C. Wan, T. Probst, L. Van Gool, and A. Yao. Crossing nets: Combining gans and vaes with a shared latent space for hand pose estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 680–689, 2017.

[36] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, p. 143–146. Association for Computing Machinery, New York, NY, USA, 2011.