

# A Modular Approach to the Embodiment of Hand Motions from Human Demonstrations

Alexander Fabisch<sup>1</sup>, Manuela Uliano<sup>2</sup>, Dennis Marschner<sup>3</sup>, Melvin Laux<sup>3</sup>, Johannes Brust<sup>4</sup>, Marco Controzzi<sup>2</sup>

**Abstract**—Manipulating objects with robotic hands is a complicated task. Not only the fingers of the hand, but also the pose of the robot’s end effector need to be coordinated. Using human demonstrations of movements is an intuitive and data-efficient way of guiding the robot’s behavior. We propose a modular framework with an automatic embodiment mapping to transfer recorded human hand motions to robotic systems. In this work, we use motion capture to record human motion. We evaluate our approach on eight challenging tasks, in which a robotic hand needs to grasp and manipulate either deformable or small and fragile objects. We test a subset of trajectories in simulation and on a real robot and the overall success rates are aligned.

## I. INTRODUCTION

Although manipulation of known objects is a well-studied field, handling deformable or small, fragile objects with human-level skill is a challenge. Behaviors for robotic hands can be generated through various approaches, e.g., planning, reinforcement learning, or imitation learning. We are interested in leveraging intuitive human knowledge to generate data for imitation learning with a complex hand. Dataset generation is difficult in this case. Kinesthetic teaching becomes tricky when a 5-finger hand and the end effector’s pose need to be controlled. Teleoperation might not exploit the full potential of the human demonstration due to restricted movement or control difficulties. We propose to use external sensors (motion capture) to track human hands and transfer their states to robotic hands. To do this, we infer the human hand’s state with a record mapping [1]. Next, we solve the correspondence problem [26], induced by kinematic differences between human and robotic hands, with an embodiment mapping [1].

Our goal is to develop a modular framework that allows us to easily replace the sensor as well as the target system (see Figure 1). For example, switching between different optical methods, e.g., motion capture and camera-based hand tracking, should be easy. For this reason, we use the MANO hand model [29], which has previously been used in camera-based hand tracking [19], as an intermediate representation of the hand’s state. The embodiment mapping should

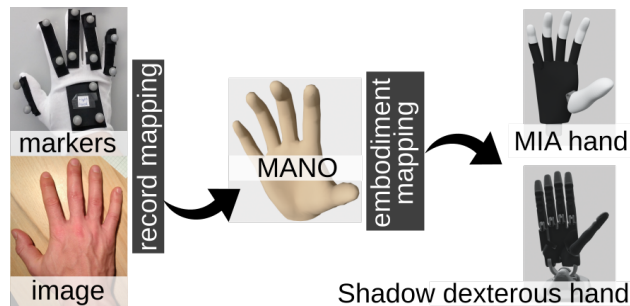


Fig. 1: Proposed approach to embodiment of hand motions.

also be configurable to handle multiple robotic hands. The implementation is available at [https://github.com/dfki-ric/hand\\_embodiment](https://github.com/dfki-ric/hand_embodiment).

## II. BACKGROUND AND RELATED WORK

### A. Motion Capture of Human Hands

Capturing human hand motions as fully articulated 3D hand poses is demanding due to the dexterity of hands and high angular velocities. Nevertheless, the task is well studied and there are numerous solutions, including optical, non-optical, and hybrid methods.

1) *Non-optical Methods*: Methods based on electromagnetic transmitters [30], [22], [4], bending [14], [31], [6] or stretch-sensors [5], [2], [15], inertial measurement units [24], [20], [9], or even exoskeletons [28] are often integrated as gloves. Caeiro-Rodríguez et al. provide a review of commercial active smart gloves [3]. These methods are suitable for real-time applications, but several problems, such as complex calibration and noisy data with drift over time, remain. Considering different hand shapes, it is not trivial to place multiple sensors perfectly on the glove without loss of accuracy or hand shape-dependent calibration methods.

2) *Optical Methods*: The continuum of optical methods ranges from estimators based on markerless, monocular color images to marker-based motion capture systems using multiple cameras. Markerless methods, mostly based on deep learning, led to groundbreaking progress in computer vision. However, various factors such as lighting conditions, image resolution, background and skin color can influence their performance. Optical markerless approaches can be divided into generative [34], [16] and discriminative methods [41], [25], [27]. Depth information can improve the accuracy of markerless optical methods. However, most methods reach their limits in everyday applications because they generalize insufficiently. Occlusion and self-occlusion when interacting with objects are problems that can be addressed with multi-view approaches [35], [36], [37], [32].

This work was supported by the European Commission under the Horizon 2020 framework program for Research and Innovation (project acronym: APRIL, project number: 870142).

<sup>1</sup> Robotics Innovation Center, DFKI GmbH, Robert-Hooke-Straße 1, D-28359 Bremen, Germany (alexander.fabisch@dfki.de)

<sup>2</sup> The BioRobotics Institute, Scuola Superiore Sant’Anna, Pisa, Italy, and with the Department of Excellence in Robotics & AI, Pisa, Scuola Superiore Sant’Anna, Italy.

<sup>3</sup> Robotics Research Group, University of Bremen

<sup>4</sup> Plan-Based Robot Control, DFKI GmbH

Optical marker-based methods (motion capture, MOCAP) are widely used in both the film industry to create realistic animations [39] and for motion analysis in sports biomechanics and rehabilitation [8]. With proper hardware and environment, professional MOCAP systems estimate hand poses more accurately than markerless methods. With a growing number of perspectives and higher camera resolutions, marker detection accuracy, as well as robustness against occlusions increase. However, such systems are expensive and of little use outside of laboratories.

### B. Human Hand Pose Models

Cobos et al. show that 24 degrees of freedom (DOF) are suitable for modeling the high kinematic complexity of human hands during manipulation [7]. Yet, no universal kinematic hand model is equally suitable for all capturing methods, and the number of measurement points varies between different hardware setups. In non-optical methods, labeled 3D joints of hand and finger key points, as well as joint angles, are commonly measured. Optical approaches usually estimate labeled 2.5D or 3D joints, but not joint angles. Mostly, labeled 3D points are assigned to a hand skeleton, which is helpful for advanced applications. However, human hand poses can also be represented as differentiable 3D hand models such as MANO [29], whose surface mesh can be fully deformed and posed. Compared to only regressing a 3D hand skeleton, this 3D hand mesh makes the method usable for computer vision and embodiment mapping.

### C. Embodiment Mapping

Embodiment mappings solve the problem of fitting movements demonstrated by a human to a robotic target system. The main challenge of this task is to deal with the differences between kinematic structures and dynamics of humans and robots. Previous works define complex objective functions that have to be solved for a complete trajectory, and focus on robotic arms [23], [17], [18] that have less variety in kinematic design than robotic hands.

We aim to design an embodiment mapping for different robotic hands and input modalities. This is achieved by using MANO [29] as an intermediate hand state representation, i.e., as an adapter between input modalities and target systems, while previous approaches only support one input modality and only work with robotic arms.

### D. Robotic Hands

We consider two robotic hands as target platforms: Prehensilia’s Mia Hand, as an example of a simple, robust robotic hand, and the Shadow Dexterous Hand of Shadow Robot Company as an example of a complex, fragile hand.<sup>1</sup>

1) *Mia Hand*: The Mia Hand is a simple, but robust robotic hand with 4 DOF that can be controlled at 20 Hz. The controllable joints are: thumb adduction/abduction (binary) and flexion, index finger flexion as well as coupled flexion of middle, ring, and little finger, which are controlled by the

<sup>1</sup>Although we mostly evaluate these two hands, the BarrettHand and the Robotiq 2F-140 gripper are already integrated in the open source release.

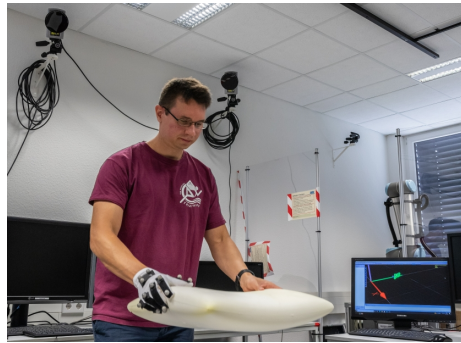


Fig. 2: Motion capture experiment.

same motor. As it is not possible to quickly switch between adduction and abduction of the thumb, we consider this joint to be fixed.

2) *Shadow Dexterous Hand*: The Shadow Dexterous Hand is complex as it has 24 DOF, of which 20 are controlled actively at 500 Hz. The last two joints of each finger (except the thumb) are coupled, such that the last joint moves when the previous one reaches the joint limit and vice versa.

## III. MODULAR RECORD AND EMBODIMENT MAPPING FOR ROBOTIC HANDS

We propose a modular framework to transfer human hand motions to robotic hands. Modularity allows us to easily adapt to new input modalities and target systems.

### A. Desiderata

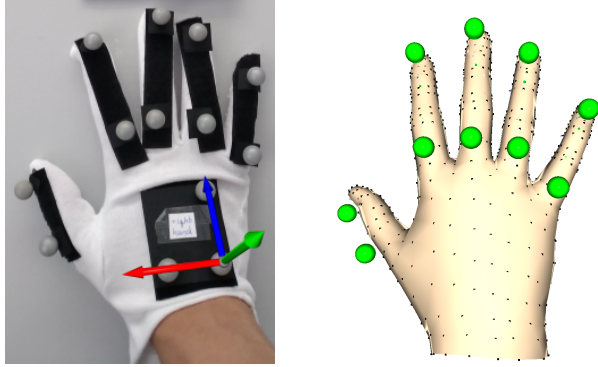
To transfer hand motions demonstrated by humans to a robot, we define a record mapping and an embodiment mapping. Our approach is designed to fulfill the following criteria:

- The approach should be adaptable to different input modalities by replacing the record mapping.
- Hence, the result of the record mapping should be a common representation of human hand states.
- The embodiment mapping should be fast enough to enable immediate transfer in teleoperation scenarios.
- The embodiment mapping should be able to adapt to the target system through configuration.

### B. Record Mapping for Motion Capture System

The objective of the record mapping is to estimate the state of the MANO model from motion capture markers. We use a Qualisys MOCAP system and a glove with 13 passive markers (see Figure 2) for the right hand. Three markers on the back of the hand and two markers per finger are used to reconstruct the pose of the hand and the configuration of each finger (see Figure 3a).

We use colors to distinguish between **estimated or measured quantities** and **configuration parameters** in formulas. To estimate the pose  $T_{\text{world,MANO}} \in SE(3)$  (read: active transformation from MANO frame to world frame) of the MANO model, we first derive the pose  $T_{\text{world,hand}} \in SE(3)$  of the hand based on three labeled markers on the back of



(a) Motion capture glove with frame defined by markers at the back of the hand.

(b) MANO model with expected marker positions indicated by green spheres.

Fig. 3: Mapping from motion capture markers to MANO.

the hand. In accordance with the two-vector representation [10], we define the hand frame orientation by the approach vector (direction from right to front hand marker) and the orientation vector (normal of the plane defined by the three markers). The origin of the hand frame can be any point in the plane of the three markers. Our frame convention is shown in Figure 3a. When we know the fixed transformation  $T_{\text{hand,MANO}} \in SE(3)$ , we compute

$$T_{\text{world,MANO}} = T_{\text{world,hand}} T_{\text{hand,MANO}}.$$

With the known pose of the MANO model, estimating the finger states boils down to solving five individual optimization problems. We compute each finger’s forward kinematics,  $f_{\beta,i,j}(\mathbf{q}) = \mathbf{p}_{i,j}$ , for the two points  $\mathbf{p}_{i,1}, \mathbf{p}_{i,2}$  (see Figure 3b), where  $\mathbf{q}_i \in \mathbb{R}^9$  are the joint angles of finger  $i \in \{1, \dots, 5\}$ . The resulting optimization problems for each finger are defined as

$$\mathbf{q}_i^* = \arg \min_{\mathbf{q}_i} \sum_{j=1}^2 \|\hat{\mathbf{p}}_{i,j} - f_{\beta,i,j}(\mathbf{q}_i)\|^2 + R(\mathbf{q}_i)$$

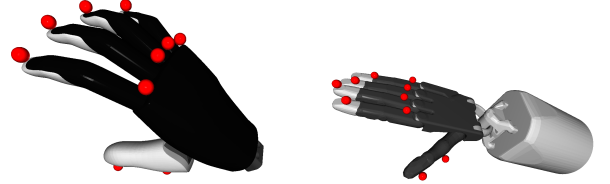
subject to  $\mathbf{q}_i^{\min} \leq \mathbf{q}_i \leq \mathbf{q}_i^{\max}$ ,

where we penalize each joint angle individually in positive and negative direction with  $R(\mathbf{q}_i) = \|\max(\mathbf{w}_{i,+} \circ \mathbf{q}_i, \mathbf{0})\|^2 + \|\min(\mathbf{w}_{i,-} \circ \mathbf{q}_i, \mathbf{0})\|^2$  with weights  $\mathbf{w}_{i,+}, \mathbf{w}_{i,-} \in \mathbb{R}^9$  ( $\circ$  is the Hadamard product and min, max are element-wise operators).  $\mathbf{q}_i^{\min}, \mathbf{q}_i^{\max} \in \mathbb{R}^9$  are lower and upper bounds for joint angles, and  $\beta \in \mathbb{R}^{10}$  are shape parameters of the MANO model.  $\hat{\mathbf{p}}_{i,j}$  are measured positions of motion capture markers. We solve these optimization problems with sequential least squares programming (SLSQP, [21]) and numerically estimated gradients.

The MANO model’s full state is defined by  $\mathbf{q}_i^* \in \mathbb{R}^9$  and  $T_{\text{world,MANO}} \in SE(3)$ , from which we can compute the corresponding marker points  $\mathbf{p}_{i,j}^* \in \mathbb{R}^3$ .

### C. Embodiment Mapping

The embodiment mapping translates MANO states to states of the target system, which is a combination of a robotic arm and hand. Assuming that poses are reachable,



(a) Model of Mia hand with expected marker positions.

(b) Model of Shadow dexterous hand with expected marker positions.

Fig. 4: Extended kinematic hand models.

the robotic hand’s pose first needs to be matched to the mesh pose, i.e., we must define  $T_{\text{robot,MANO}} \in SE(3)$ .

Next, the individual finger configurations are optimized to be as close as possible to the MANO mesh. Without real-time constraints, the ideal solution is to define an objective function to either maximize the overlap between the volumes or to minimize the distance between the inner surfaces of MANO’s fingers and the fingers of the robotic hand. With the intention to be able to transfer motions in real-time, we propose a simplified approach. We define points with respect to the links of the robotic hands (see Figure 4) and minimize the distance to their corresponding virtual markers on the MANO mesh (see Figure 3b), for which the positions are known from the record mapping.

Thus, the optimization of finger joints reduces to an inverse kinematics problem, in which only the distance between two pairs of points per finger  $i$  is minimized:

$$\mathbf{r}_i^* = \arg \min_{\mathbf{r}_i} \sum_{j=1}^2 \|f_{\beta,i,j}(\mathbf{q}_i^*) - g_{i,j}(\mathbf{r}_i)\|^2,$$

subject to  $\mathbf{r}_i^{\min} \leq \mathbf{r}_i \leq \mathbf{r}_i^{\max}$ .  $f_{\beta,i,j}(\mathbf{q}_i^*)$  is known from record mapping and  $g_{i,j}(\mathbf{r}_i)$  is the corresponding forward kinematics function for the robotic hand with the joint angles  $\mathbf{r}_i \in \mathbb{R}^{N_i}$  and limits  $\mathbf{r}_i^{\min}, \mathbf{r}_i^{\max} \in \mathbb{R}^{N_i}$ . The number of optimized joints  $N_i \in \mathbb{N}$  depends on the target system, as e.g., the Mia hand’s index finger is controlled by a single motor while the Shadow dexterous hand uses three motors to control the index finger. The optimization problem is solved by SLSQP.

### D. Configuration

While we assume well-defined kinematics, it is necessary to configure certain parameters of the record and embodiment mapping. For this work, these parameters were configured manually, however, this could be partially automated. For instance, a black-box optimizer could set the shape parameters for the MANO model to fit motion capture markers.

For the record mapping, we need to configure:

- $T_{\text{hand,MANO}} \in SE(3)$ : transformation between MANO base and hand coordinate frame defined by three motion capture markers at the back of the hand
- $\beta \in \mathbb{R}^{10}$ : shape parameters of MANO
- $\mathbf{w}_{i,+}, \mathbf{w}_{i,-} \in \mathbb{R}^9$ : weights to penalize each joint angle individually in both directions

No.	Task	Variations	Demonstrations
1	Grasp insole	from front or back	213
2	Insert insole	-	12
3	Grasp small pillow	from four sides	224
4	Grasp big pillow	from four sides	130
5	Grasp electronic component	from all sides	55
6	Assemble electronic components	from all directions	54
7	Flip pages	-	38
8	Insert passport in box	-	37
Total			763

TABLE I: Overview of datasets used for evaluation.



Fig. 5: Objects used to record datasets. Left to right and top to bottom: insole with markers, insole and bag, small pillow with markers, open passport, passport and box, electronic components with markers.

- $q_i^{\min}, q_i^{\max} \in \mathbb{R}^9$ : joints' lower and upper bounds

For the embodiment mapping, we need to configure

- $T_{\text{robot}, \text{MANO}} \in SE(3)$ : transformation between the MANO mesh's and the robotic hand's bases
- expected marker positions (see Figure 4) with respect to corresponding frames in the hand's kinematic tree

## IV. EVALUATION

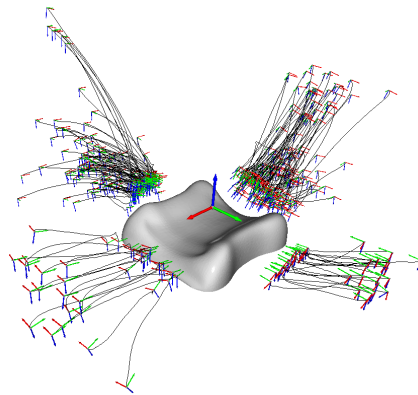
### A. Research Question

It has been shown that the state of the MANO model can be obtained from RGB images [19]. Our goal is to evaluate whether

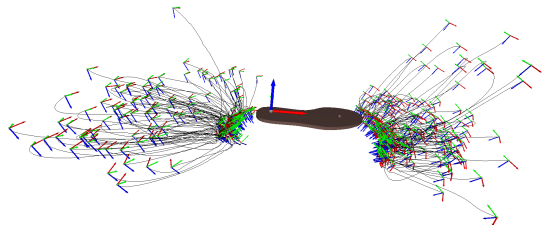
- 1) It is possible to obtain the MANO representation from motion capture data.
- 2) The embodiment mapping can be adapted to both robotic hands.
- 3) The embodiment mapping obtains plausible configurations of the robotic hand even when the target system has less DOF than a human hand.
- 4) The transferred trajectories result in physically verified useful behaviors of the target systems.
- 5) Both record and embodiment mapping can be executed at a frequency suitable for teleoperation.

### B. Datasets

To evaluate the hand embodiment mapping, we recorded demonstrations of multiple tasks and variations of these with a Qualisys MOCAP system. Table I describes the tasks and



(a) Dataset of 224 grasps for a small pillow.



(b) Dataset of 213 grasps for an insole.

Fig. 6: Two of the datasets used for evaluation. Object-relative trajectories of the end effector are represented by lines and coordinate frames that indicate the orientation at the beginning and end. The large coordinate frames in the middle define object poses.

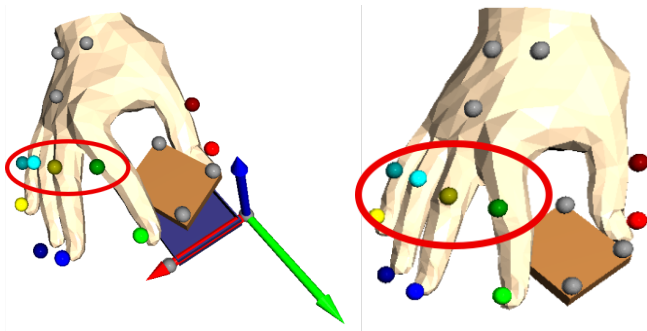
reports the number of demonstrations for each of these.<sup>2</sup> Objects that were used during these experiments are shown in Figure 5 and visualizations of the end-effector trajectories in Figure 6.

### C. Estimation of MANO State (Qualitative Evaluation)

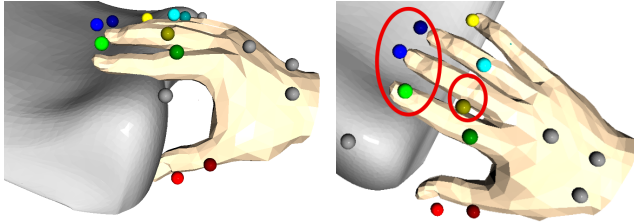
Marker positions were tracked with an error of about 1 mm. We can also use them to evaluate the quality of the estimated MANO states. Figure 7 shows exemplary measurements of the motion capture markers with corresponding estimations of the MANO model by the record mapping from marker positions. Differences between the MANO state and the actual hand mainly stem from inaccuracies of the MANO configuration. In particular, shape and the placement of the three markers at the back of the MANO model are different. Marker placement also varies between experiments and even within individual recordings.

Obvious differences between the estimated MANO state and the actual hand state can be seen, e.g., in Figure 7a (red ellipses): while the marker positions are closer to the metacarpophalangeal joint of the real hand (see Figure 3a), they are closer to the proximal interphalangeal joint of the estimated MANO state. Furthermore, we can see in Figure 7b (red ellipses) that the lengths of the fingers do not always match the corresponding marker positions. In the

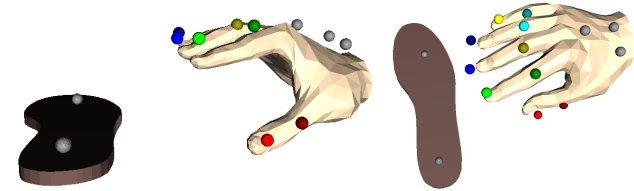
<sup>2</sup>Only one subject was recorded because of COVID-19. We argue that this is sufficient since parameters of the record mapping are tuned manually.



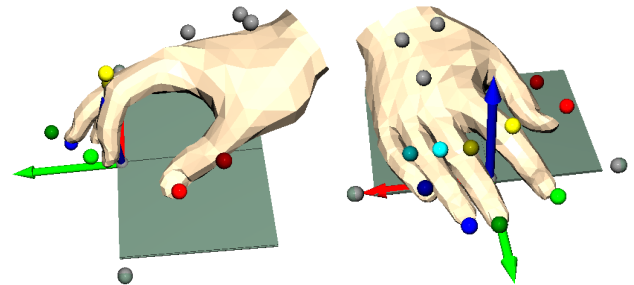
(a) Grasping and assembling electronic components.



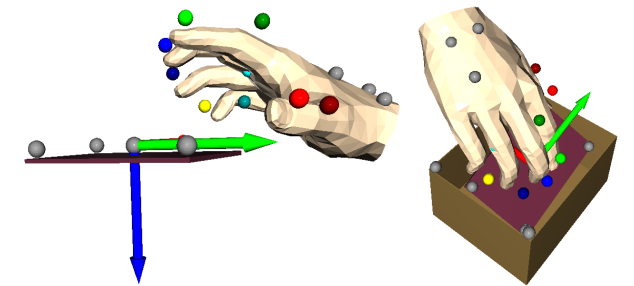
(b) Grasping a small pillow.



(c) Grasping an insole.



(d) Flipping the page of a passport.



(e) Grasping a passport and putting it in a box.

Fig. 7: Exemplary configurations of MANO mesh after record mapping and corresponding motion capture markers. Simplified meshes that illustrate the position of the manipulated objects are displayed with the markers that we used to track their pose. For some objects we also see the object frame. Illustrations were made with Open3D [40].

same example, the marker close to the metacarpophalangeal joint of the middle finger is laterally shifted, which was due to the marker not being perfectly aligned at the center of the finger. Nevertheless, we can see in Figure 7 that the estimated MANO states are generally plausible explanations of the measured marker positions.

#### D. Adaptability to Robotic Hand (Qualitative Evaluation)

Figure 8 shows the result of an interactive embodiment mapping. A GUI application was used to set the 48 joint parameters of the MANO model. The embodiment mapping determines joint angles of the robotic hand. Both the MANO mesh and the configuration of the robotic hand are visualized.

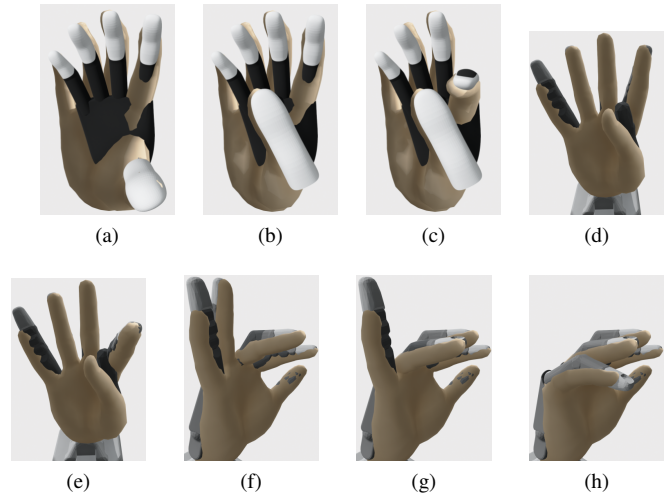


Fig. 8: Interactive embodiment mapping. MANO state and robotic hand after embodiment mapping are displayed together. This visualization is based on Open3D's visualizer [40] and pytransform3d [13].

Figure 8 shows exemplary configurations of the Mia hand (a – c) and configurations of the Shadow dexterous hand (d – h). There are differences between the MANO mesh and the robotic hand that the embodiment mapping cannot compensate for: the Mia hand is slightly smaller than the MANO mesh so that the little finger cannot be aligned perfectly, and the little finger of the Shadow dexterous hand is longer than the one of the MANO mesh. The Mia hand has only 4 DOF, which results in a less accurate embodiment, in particular when the middle finger, ring finger, and little finger have a different flexion as these move jointly in the Mia hand. There are also differences that occur due to an inadequate objective: the Shadow dexterous hand is able to minimize the positional difference between the finger tips without having the correct orientation (see Figure 8d) and as long as the tip positions are reached it does not matter whether the joint angles are similar (e.g., see Figure 8h). As it will become apparent in Sections IV-E and IV-H, the last point is the price that we pay to for real-time control of a robotic hand.



Fig. 9: Contact surfaces of the MANO model and the robotic hands are marked in red color.

Hand	Thumb	Index	Middle	Ring	Little
Mia hand	12.2	8.3	17.4	23.3	30.2
Shadow dexterous hand	5.2	6.3	5.6	5.9	9.8
Robotiq 2F-140	28.6	13.4	-	-	-

TABLE II: Mean average distance (unit: mm) of contact surfaces. One frame per demonstration (763) was selected. A simple Robotiq gripper with 1 DOF is used as baseline.

### E. Similarity Between MANO and Robotic Hand

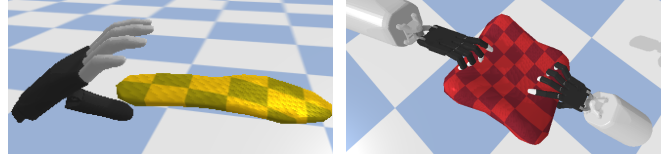
We compare inner surfaces of fingers of the robotic hands to the MANO mesh to evaluate the embodiment. Hence, we define the contact surfaces of the MANO model and each robotic hand for each finger that we compare (see Figure 9).

For evaluation we draw 100 points from the contact surface of each finger of the robotic hand by Poisson disk sampling [38], compute the minimum distance to the closest triangle of the corresponding surface of the MANO mesh per sample, and average these minimum distances over all samples per finger. More precisely, we compute  $\frac{1}{N} \sum_{i=1}^N \min_{j \in \{1, \dots, M\}} d(\mathbf{p}_i, T_j)$ , where  $d(\mathbf{p}, T)$  is the distance between a point and a triangle [12],  $\mathbf{p}_i$  are points on the contact surface of the robotic finger,  $T_j$  are triangles of the corresponding finger of MANO,  $N$  is the number of samples from the robotic finger, and  $M$  is the number of triangles on the contact surface of MANO. The result is an average distance between the two surfaces. Since the computation is considerably slower than the embodiment, we do it for selected cases only.

Table II shows that more DOF enable the embodiment mapping to fit desired configurations more closely. The Mia hand, e.g., has problems with fitting the middle, ring and little fingers because they are controlled by the same motor.

### F. Transfer to Simulation

We use PyBullet [11], one of the few physics engines that support robots and deformable objects, to verify physical plausibility of transferred motions. Since setting up realistic simulation environments and modeling deformable objects that have complex shapes is difficult, we focus on the tasks of grasping an insole and a small pillow. We model them as homogeneous objects with the stable Neo-Hookean model [33] for hyperelastic material. For the insole we set Young’s modulus to  $E = 100$  kPa and Poisson’s ratio to  $\nu = 0.2$ . For the pillow we set  $E = 10$  kPa and  $\nu = 0.2$ .



(a) Insole and Mia hand. (b) Pillow and Shadow hands.

Fig. 10: Simulation with floating hands.

Object	Samples	Hand	Success Rate
Insole	213	Mia	71.4%
		Shadow	40.4%
Small pillow	224	Mia	92.8%
		Shadow	65.2%

TABLE III: Success rates of simulated grasps.

We test whether the object can be held after the execution of each grasp by simulating the effect of gravity. The objects float initially (see Figure 10). After each completed grasp, we evaluate its success by allowing the object to fall from gravitational force. We continue the simulation for two seconds and measure whether the object is still in the hand. To exclude problems of reachability, we simulate only floating hands without a robotic arm. Since the pillow is much larger than both hands, the best strategy is to grasp it with two hands. Hence, we execute the same demonstrated grasp with two hands, where one trajectory is rotated 180 degrees around the axis pointing up in the middle of the pillow, which works because the pillow is symmetric. Otherwise a second human hand would have to be recorded.

Table III shows the success rate of the embodiment mapping for each combination of task and hand. Considering morphological differences between the human hand and the robotic hands 100% success rate is hardly achievable. Despite resembling MANO states more faithfully, the Shadow hand does not perform better than the Mia hand in these tasks. We attribute this to the fact that it is not necessary to have many DOF to match the recorded human cylindrical and pinch grasps. In fact, the geometry of the Mia hand is better suited to grasp these two objects firmly than the Shadow hand, mainly due to its big thumb. Note that we expect grasping to work better when force sensors are used as feedback. Thus, it is best to use the transferred motion in combination with a controller that could, e.g., be generated through reinforcement learning. Modeling contact and friction of deformable objects in simulation is difficult. Hence, we perform experiments on the real system to check whether we can trust results from simulation.

### G. Transfer to Real System

1) *Methods:* The trajectories generated by the embodiment were tested in a real robot. The set-up comprises a robot arm (UR10e, Universal Robots), a 6-axis force-torque sensor (HEX-E v2, OnRobot) mounted at the wrist of the robot arm, and an anthropomorphic artificial hand (MIA hand, Prensilia). The target object is a deformable insole

bending at the edge of a table and the grasp point is outside of the table. During the tests, the insole was positioned in the same location by means of a mask. The arm and the hand were controlled at a frequency of 100 Hz and 20 Hz, respectively. The data from the embodiment mapping were used to control the robot via a multi-node ROS environment. The experiment included a subset of 80 trajectories differing for grasp location (i.e., 40 tip and 40 heel), but characterized by the same grasp type (i.e., cylindrical). The experiment comprised of two experimental conditions. The first (Coordinated Trajectories - CT), is aimed at assessing the capacity of the embodiment to successfully transfer coordinated motions. Thus, in this session, the reaching motion of the arm and the grasping action of the hand were controlled in a coordinated fashion as computed by the embodiment. The second (Sequential Trajectories - ST) sought to assess the success rate when the reaching motion of the arm is accomplished before the grasping action of the hand. For each trial, the robot executed the trajectory of the arm and the joint trajectories of fingers to grasp the insole. At the end, the robot moved toward the human operator through a predefined trajectory. The grasp was judged successful if the object did not fall during the lifting and transporting phases (without changing the object orientation as for the simulation). Finally, if the grasping action of the robot succeeded, the human operator pulled the object out of the hand along the direction of the long fingers. In this phase, we recorded the magnitude of the force at the wrist of the robot and used this data to indirectly evaluate the stability of the grasp.

2) *Results and discussion:* Results of the experiments are summarized in Table IV. Overall, the success rate is 83.8% (67 out of 80 trajectories) for the CT condition, and 62.5% (50 out of 80 trajectories) for the ST condition. This trend is also confirmed by looking at the performances for the different grasp locations. These results show that the embodiment mapping preserves the benefits of the human motor coordination in the action of grasping. Among the different grasp locations, the heel has the highest success rate (95% and 75% for the CT and ST conditions, respectively). Successful trajectories can lead to stable or slightly stable grasps, and we used the force recorded during the pull-out phase to discriminate among these two classes. A grasp was judged stable if the maximum of the magnitude of the force is greater than a threshold of 2.8 N (this value was set based on the performance of the sensor used). Overall, the success rate of stable grasps is 72.5% (58 out of 80 trajectories) for the CT condition. This result is aligned with the output of the simulation, being 70% considering this subset of 80 trajectories (Table IV).

#### H. Real-Time Control Capabilities

One intended use case of the embodiment mapping is to enable real-time control of a robotic arm and hand through a motion capture system. A limiting factor is the frequency at which we receive hand states from the motion capture system, which is 100 Hz. We must generate commands for the robotic hand from motion capture with a similar

		Success rate	Success (trajectories) #	
Coordinated Trajectories	Tip	All	72.5%	29 (40)
		Stable	62.5%	25 (40)
		Simulation	62.5%	25 (40)
	Heel	All	95.0%	38 (40)
		Stable	82.5%	33 (40)
		Simulation	77.5%	31 (40)
Overall	All	83.8%	67 (80)	
	Stable	72.5%	58 (80)	
	Simulation	70.0%	56 (80)	
Sequential Trajectories	Tip	All	50.0%	20 (40)
		Stable	42.5%	17 (40)
	Heel	All	75.0%	30 (40)
		Stable	62.5%	25 (40)
	Overall	All	62.5%	50 (80)
		Stable	53.8%	43 (80)

TABLE IV: Success rates and stability of the grasps for the set of 80 trajectories executed in coordinated and sequential fashion.

Task no.	Frames	Record/Embodiment Mapping Frequency in Hz					
		MANO		Mia		Shadow	
		mean	min	mean	min	mean	min
1	42,285	99.7	5.1	228.0	15.0	178.7	0.7
3	31,020	65.8	5.9	261.1	102.2	136.7	9.0
6	16,984	95.2	7.7	293.9	123.1	180.7	10.5
7	15,484	63.4	9.1	263.1	108.5	128.0	0.7

TABLE V: Evaluation of speed. Results are statistics of each frame of each demonstration of the task (only tasks with >10,000 frames). Computations are done by one core of an AMD Ryzen 7 2700 CPU. Task numbers refer to Table I.

frequency. Table V shows the frequencies at which we are able to compute record and embodiment mapping. While the lowest frequencies for both mappings prevent real time control even with a control frequency of 20 Hz for the Mia hand, the average frequency of the embodiment mapping is well above the frequency at which the motion capture system provides measurements. The record mapping is often too slow for 100 Hz. Nevertheless, it is possible to split hand pose estimation, which can be done at a high frequency, and estimation of finger configuration, which does not need to be done at a high frequency to control the Mia hand.

## V. CONCLUSIONS

We show that MANO states can be obtained from MOCAP in addition to the usual approach with one camera. Using MOCAP allows for more complex, natural motions than kinesthetic teaching and occlusions are less likely than with a single camera. Furthermore, we introduce a modular framework to transfer human hand motions to two robotic hands with varying complexity through a configurable embodiment mapping that is fast enough for complex robotic hands. Even without feedback, most embodied trajectories could successfully solve the task of grasping simulated insoles and pillows. Real results are aligned with simulation. Experiments also show that coordination between hand pose

and finger movements is more effective than sequential execution, which emphasizes the relevance of our approach. However, results could be improved with, e.g., reinforcement learning. On the one hand this would considerably reduce the necessary exploration for reinforcement learning and on the other hand it would make transferred motions more robust.

### ETHICS APPROVAL

Experimental protocols were approved by the ethics committee of the University of Bremen. Written informed consent was obtained from all participants.

### ACKNOWLEDGMENT

We thank Oscar Lima, Andrea Burani, and Francesca Cini for the URDF of the Mia hand and Lisa Gutzeit for her feedback on the manuscript. The motion capture setup was developed in collaboration with Lisa Gutzeit, supported by a grant from the German Federal Ministry for Economic Affairs and Energy (BMW, FKZ 50 RA 2023).

### REFERENCES

- [1] B. D. Argall, S. Chernova, M. Veloso, and B. Browning. A survey of robot learning from demonstration. *Robotics and Autonomous Systems*, 57(5):469–483, 2009.
- [2] A. Atalay, V. Sanchez, O. Atalay, D. Vogt, F. Haufe, R. Wood, and C. Walsh. Batch fabrication of customizable silicone-textile composite capacitive strain sensors for human motion tracking. *Advanced Materials Technologies*, 2, 07 2017.
- [3] M. Caeiro-Rodríguez, I. Otero-González, F. A. Mikic-Fonte, and M. Llamas-Nistal. A systematic review of commercial smart gloves: Current status and applications. *Sensors*, 21(8), 2021.
- [4] K.-Y. Chen, S. Patel, and S. Keller. Finexus: Tracking precise motions of multiple fingertips using magnetic sensing. pages 1504–1514, 2016.
- [5] J.-B. Chossat, Y. Tao, V. Duchaine, and Y.-L. Park. Wearable soft artificial skin for hand motion detection with embedded microfluidic strain sensing. In *ICRA*, pages 2568–2573, 2015.
- [6] S. Ciotti, E. Battaglia, N. Carbonaro, A. Bicchi, A. Tognetti, and M. Bianchi. A synergy-based optimally designed sensing glove for functional grasp recognition. *Sensors (Basel, Switzerland)*, 16, 2016.
- [7] S. Cobos, M. Ferre, M.A. Uran, J. Ortego, and C. Peña Cortés. Efficient human hand kinematics for manipulation tasks. pages 2246–2251, 2008.
- [8] S. L. Colyer, M. Evans, D. P. Cosker, and A. I. T. Salo. A review of the evolution of vision-based motion analysis and the integration of advanced computer vision methods towards developing a markerless system. *Sports Medicine - Open*, 4(1):24, 2018.
- [9] J. Connolly, B. O’Flynn, J. Torres Sanchez, J. Condell, K. Curran, P. Gardiner, and B. Downes. Integrated smart glove for hand motion monitoring. 01 2015.
- [10] P. Corke. *Robotics, Vision and Control*. Springer, 2017.
- [11] E. Coumans and Y. Bai. PyBullet, a python module for physics simulation for games, robotics and machine learning. <http://pybullet.org>, 2016–2022.
- [12] C. Ericson. *Real-Time Collision Detection*. CRC Press, 2004.
- [13] A. Fabisch. pytransform3d: 3d transformations for python. *Journal of Open Source Software*, 4(33):1159, 2019.
- [14] R. Gentner and J. Classen. Development and evaluation of a low-cost sensor glove for assessment of human finger movements in neurophysiological settings. *J. Neurosci. Methods*, 178:138–147, 2009.
- [15] O. Glauser, S. Wu, D. Panozzo, O. Hilliges, and O. Sorkine-Hornung. Interactive hand pose estimation using a stretch-sensing soft glove. *ACM Trans. Graph.*, 38(4), jul 2019.
- [16] J. Gu, Z. Wang, W. Ouyang, W. Zhang, J. Li, and L. Zhuo. 3d hand pose estimation with disentangled cross-modal latent space. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, March 2020.
- [17] L. Gutzeit, A. Fabisch, M. Otto, J. H. Metzen, J. Hansen, F. Kirchner, and E. A. Kirchner. The besman learning platform for automated robot skill learning. *Frontiers in Robotics and AI*, 5:43, 2018.
- [18] L. Gutzeit, A. Fabisch, C. Petzoldt, H. Wiese, and F. Kirchner. Automated robot skill learning from demonstration for various robot systems. In Christoph Benzmüller and Heiner Stuckenschmidt, editors, *KI: Advances in Artificial Intelligence*, pages 168–181. Springer, 2019.
- [19] Y. Hasson, G. Varol, D. Tzionas, I. Kalevatykh, M. J. Black, I. Laptev, and C. Schmid. Learning joint reconstruction of hands and manipulated objects. In *CVPR*, 2019.
- [20] P.-C. Hsiao, S.-Y. Yang, B.-S. Lin, I.-J. Lee, and W. Chou. Data glove embedded with 9-axis imu and force sensing sensors for evaluation of hand function. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4631–4634, 2015.
- [21] D. Kraft. A software package for sequential quadratic programming. Technical Report DFVLR-FB 88-28, DLR German Aerospace Center – Institute for Flight Mechanics, Köln, Germany, 1988.
- [22] Y. Ma, Z.-H. Mao, W. Jia, C. Li, J. Yang, and M. Sun. Magnetic hand tracking for human-computer interface. *IEEE Transactions on Magnetics*, 47(5):970–973, 2011.
- [23] G. Maeda, M. Ewerton, D. Koert, and J. Peters. Acquiring and generalizing the embodiment mapping from human observations to robot skills. *IEEE Robot. and Autom. Lett.*, 1(2):784–791, 2016.
- [24] T. Mańkowski, J. Tomczyński, and P. Kaczmarek. Cie-dataglove, a multi-imu system for hand posture tracking. pages 268–276, 03 2017.
- [25] F. Mueller, F. Bernard, O. Sotnychenko, D. Mehta, S. Sridhar, D. Casas, and C. Theobalt. Generated hands for real-time 3d hand tracking from monocular rgb. In *CVPR*, June 2018.
- [26] C. L. Nehaniv and K. Dautenhahn. *The Correspondence Problem*, pages 41–61. MIT Press, Cambridge, MA, USA, 2002.
- [27] P. Panteleris, I. Oikonomidis, and A. A. Argyros. Using a single RGB frame for real time 3d hand pose estimation in the wild. *CoRR*, abs/1712.03866, 2017.
- [28] A. Pereira, G. Stillfried, T. Baker, A. Schmidt, A. Maier, B. Pleintinger, Z. Chen, T. Hulin, and N. Y. Lii. Reconstructing human hand pose and configuration using a fixed-base exoskeleton. In *ICRA*, pages 3514–3520, 2019.
- [29] 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), November 2017.
- [30] H.-M. Shen, C. Lian, X.-W. Wu, F. Bian, P. Yu, and G. Yang. Full-pose estimation using inertial and magnetic sensor fusion in structured magnetic field for hand motion tracking. *Measurement*, 170:108697, 2021.
- [31] Z. Shen, J. Yi, X. Li, L. H. P. Mark, Y. Hu, and Z. Wang. A soft stretchable bending sensor and data glove applications. In *IEEE International Conference on Real-time Computing and Robotics (RCAR)*, pages 88–93, 2016.
- [32] T. Simon, H. Joo, I. Matthews, and Y. Sheikh. Hand keypoint detection in single images using multiview bootstrapping, 2017.
- [33] B. Smith, F. De Goes, and T. Kim. Stable neo-hookean flesh simulation. *ACM Trans. Graph.*, 37(2), 2018.
- [34] A. Spurr, J. Song, S. Park, and O. Hilliges. Cross-modal deep variational hand pose estimation. *CoRR*, abs/1803.11404, 2018.
- [35] S. Sridhar, A. Oulasvirta, and C. Theobalt. Interactive markerless articulated hand motion tracking using rgb and depth data. In *ICCV*, 2013.
- [36] D. Tzionas, L. Ballan, A. Srikantha, P. Aponte, M. Pollefeys, and J. Gall. Capturing hands in action using discriminative salient points and physics simulation. *CoRR*, abs/1506.02178, 2015.
- [37] R. Wang, S. Paris, and J. Popović. 6d hands: Markerless hand-tracking for computer aided design. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, UIST ’11*, page 549–558, New York, NY, USA, 2011. ACM.
- [38] C. Yuksel. Sample elimination for generating poisson disk sample sets. *Computer Graphics Forum (Proceedings of EUROGRAPHICS)*, 34(2):25–32, 2015.
- [39] M. Y. Zhang. Application of performance motion capture technology in film and television performance animation. In *Instruments, Measurement, Electronics and Information Engineering*, volume 347 of *Applied Mechanics and Materials*, pages 2781–2784. Trans Tech Publications Ltd, 10 2013.
- [40] Q.-Y. Zhou, J. Park, and V. Koltun. Open3D: A modern library for 3D data processing, 2018.
- [41] C. Zimmermann and T. Brox. Learning to estimate 3d hand pose from single rgb images. In *ICCV*, 2017. <https://arxiv.org/abs/1705.01389>.