

High Fidelity Capture, Reconstruction, and Transfer of Human Demonstrations for Robot-Assisted Bathing

Arjun S. Lakshmipathy¹, Jonathan P. King¹, Ethan Zuo^{1*}, Rohit Satishkumar^{1*}, Hongyi Chen¹, Jeffrey Ichnowski¹, Dan Ding², Zackory Erickson¹, and Nancy S. Pollard¹

Abstract—Despite the demand for robots in high-value clinical tasks like bathing, contemporary systems still lack the safety and reliability required for complex, sustained physical interaction with humans. A key challenge hindering the development of such systems is that collecting, understanding, and effectively transferring highly dynamic, contact-rich human bathing demonstrations is difficult, even with modern motion and tactile sensing equipment. We present a straightforward, but effective framework for doing so with high fidelity by utilizing contact regions as a key processing primitive. We use our framework to build a dataset of bathing demonstrations performed by trained clinicians on human subjects. We then use this dataset to design and control an arm-mounted dexterous soft hand to perform bathing tasks on a mannequin using open- and closed-loop strategies. Our dataset is the first to provide high quality synchronized motion, shape, contact, and force during sustained, contact-rich human-human interaction, and our transfer strategies demonstrate effective use of these data across multiple levels of the robotics stack. All relevant materials will be publicly released to enable further advancements in physical human-robot interaction (pHRI) research.

I. INTRODUCTION

Bathing is an essential activity of daily living and important component of long-term personal hygiene [1]. The Katz scale [2] identifies the ability to bathe as one of six major categories that quantify an individual’s level of independence; consequently, losing this ability significantly impacts quality of life. As the number of Americans with bathing-related impairments rises and the pool of trained caregivers shrinks [3], the demand for bathing assistance is rapidly outstripping the available supply of care.

Robot-assisted bathing solutions have the potential to significantly expand access to care while reducing existing caregiver burden [4, 5]. However, building systems that can safely and reliably perform the sustained, highly dynamic, and contact-rich human interactions inherent to bathing is technically challenging. While several impressive systems have emerged recently [6, 7, 8], they generally rely on idealized definitions of bathing ranging from best practice guidelines [8] to coarse-grained performance metrics (e.g. debris removal percentages [9, 7]) rather than real data from human demonstrations. Consequently, many important details such as realistic hand

movements and shape adjustments across different body segments, reactive corrections in response to patient movement, the evolution of contact regions between the hand and body over time, and forces applied over those regions are not considered. To date, there are no comprehensive data that capture how caregivers actually bathe people.

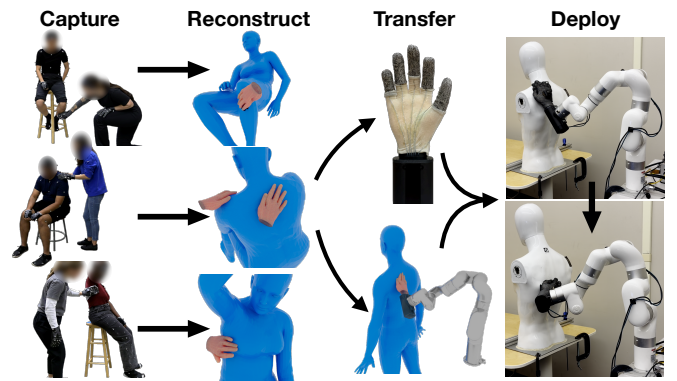


Fig. 1: Our contributions enable straightforward, yet high quality reconstructions of human bathing demonstrations along with strategies for robot skill transfer.

This paper presents a straightforward, but effective approach for high fidelity capture and reconstruction of human bathing demonstrations, as well as techniques that utilize this information to facilitate the design and control of robot systems. Our capture pipeline takes full advantage of modern motion and tactile capture systems, but crucially provides solutions and guidance to address issues that come with reconstructing fine-grained physical interaction. Central to our techniques is the use of contact regions as a key processing primitive, which we demonstrate naturally addresses many pain points in the pipeline. We make the following contributions:

- A high quality reconstruction pipeline for time-synchronized motion, shape, contact, and force of real human-assisted bathing demonstrations
- A dataset of 128 captures and 257,000 total frames sourced across three clinicians and six human subjects
- Design and control strategies for an anthropomorphic tendon-driven soft hand using our dataset
- A retargeting pipeline to transfer demonstrations to a composite arm-hand system

¹School of Computer Science, Carnegie Mellon University

²School of Health and Rehabilitation Sciences, University of Pittsburgh

*Denotes equal contribution

Our contributions enable visualization, understanding, and translation of human-assisted bathing demonstrations at an unprecedented level of detail, which we anticipate will be of considerable interest to the clinical and pHRI communities.

II. RELATED WORK

We discuss related work spanning robot-assisted bathing systems, tactile-augmented human motion capture and transfer, and contract-driven strategies for grasping and manipulation.

A. Robot-Assisted Bathing Systems

Multiple robot-assisted bathing platforms have previously been proposed, ranging from fixed installations to mobile manipulators. Fixed installations [10, 11] are often custom-engineered for specific robotic platforms. As a result, bathing tasks and motion planning must be developed from scratch for each system and location, precluding the direct transfer of skills from human demonstrations. Alternative approaches utilize mobile manipulators [12, 8, 7, 6] to bypass location constraints, though the specific methodologies employed vary. Liu et. al. [7] propose a unique wearable design optimized for debris cleanup. King et. al. [12], Erickson et. al. [6], and most recently Madan et. al. [8] use more traditional designs and instead focus their efforts on perception challenges inherent in bathing tasks. However, these works prioritize the development of robot-specific policies over the capture and transfer of human demonstrations. We focus on the latter, investigating how human skills can be effectively captured, reconstructed, and translated to robotic platforms.

B. Tactile-Augmented Human Motion Capture and Transfer

Recent research has increasingly focused on jointly capturing motion and force information from human demonstrations, as well as transferring skills to robots. A wide array of sensing modalities have been investigated to date, ranging from force-sensitive resistive [13, 14] and capacitive arrays [15] to more specialized approaches such as embedded-camera-based [16], magnetic [17, 18], acoustic [19], and Fiber Bragg Grating sensors [20]. Skill transfer strategies, which are predominantly learning-based, include using demonstrations to train transformers [17] or condition diffusion policies [18]. Our work extends this literature to pHRI, which is a challenging domain characterized by sustained, predominantly sliding contact-rich interactions across the imprecise and shifting geometry of the human body. Concretely, we utilize capacitive sensing for our captures but focus more on data processing and straightforward methods of transfer at multiple layers of the robotics stack.

C. Contact-Driven Strategies for Grasping and Manipulation

Contact information is fundamental to grasping and manipulation; it is frequently leveraged as a loss term for pose optimization [21, 22, 23], a constraint for training generative models [24, 25], a prior for retargeting human trajectories [26, 27, 28], and a primitive for motion planning [29, 30]. Modern high quality human motion and interaction datasets [31, 32, 33, 34], many of which serve as the backbones and testbeds

of manipulation research efforts, are notably prioritizing the inclusion of detailed contact information. We take these efforts one step further by showing that high fidelity bathing reconstruction and transfer are *critically dependent* on the effective use of contact areas as a processing primitive.

III. METHODS

Figure 2 provides an overview of our capture, reconstruction, and transfer pipeline. We detail each component below.

A. Data Capture

We collected demonstrations from three trained clinicians and six subjects. Each clinician was randomly assigned to two subjects. Subjects were required to be of at least 18 years of age and able-bodied, while clinicians were required to have prior experience with assisted bathing. Subjects were outfitted with 52 optical markers for motion tracking and encouraged, but not required, to wear dark summer clothing to minimize optical reflectance while exposing as many extremities as possible for direct skin contact. Clinicians were instructed to wear a pair of commercially available tactile sensing gloves [15] and perform bathing motions on the subject’s skin, or alternatively clothing if skin exposure was not possible, using a hypoallergenic wet wipe. The dorsal side of each glove was outfitted with 23 optical markers, while the palmar side was outfitted with a layer of adhesive-backed water-resistant grip material [37] to protect the sensor electronics from incidental moisture damage from the wipe. The grip material was laser cut into a “tabbed” pattern designed to maximize surface coverage with minimal impact to range-of-motion (ROM). Notably, we found that the grip material provided the best compromise between water resistance, adhesion, and deformability in comparison to other material candidates (e.g. vinyl, silicone). Figures 3 and 4 illustrate our equipment.

The sensory array of each glove consists of 65 distinct “taxels”, or pressure-sensitive pads, distributed throughout the palmar and digital surfaces. The sensors measure changes in pressure via differential capacitance [38] relative to a baseline taken at the start of each collection. We assume all forces are normal to each taxel.

Subjects were positioned upright on a stool surrounded by a set of 20 Vicon Vantage-V16 cameras [39]. Clinicians were asked to bathe subjects in the same way as they would a patient in a medical or assisted-care facility. Markers inhibiting bathing motions of specific body parts were temporarily removed during captures of those parts.

Each capture consisted of the tracked 3D positions of all subject markers and glove optical markers, as well as pressure readings from all 65 taxels per glove. Tactile and marker data were streamed at 10 and 120 Hz respectively and synchronized using Vicon timecodes. All subjects and clinicians were also asked to perform ROM tests prior to their first capture. We ensure all markers are labeled and have unbroken trajectories.

B. Shape and Motion Reconstruction

We convert tracked marker positions and force measurements into complete reconstructions of shape, motion, and

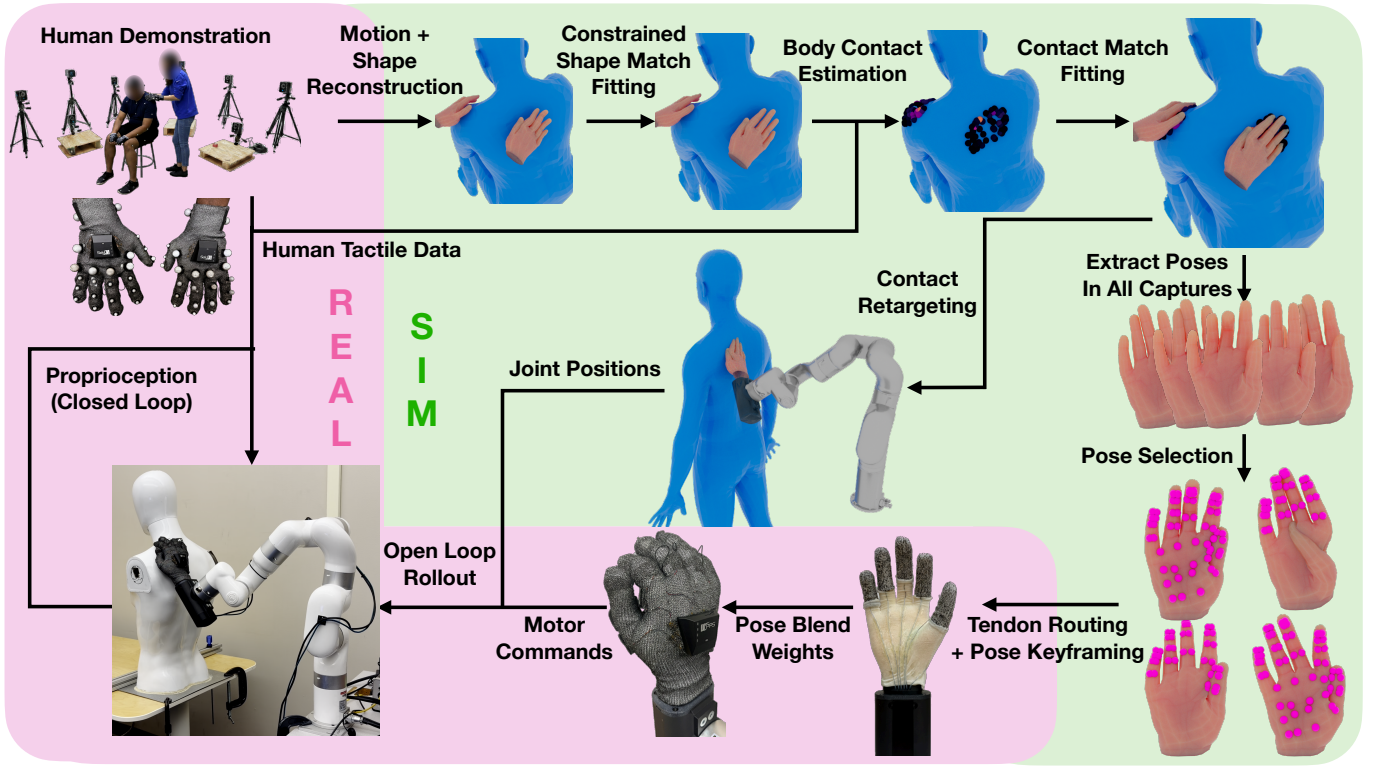


Fig. 2: Overview of our pipeline. Starting with captures of motion and tactile forces from human bathing demonstrations, we digitally reconstruct time-synchronized data across shape, movement, and contact force with high accuracy. We then utilize these reconstructions to both design and control an anthropomorphic soft hand, as well as retarget results to a robot arm. We demonstrate the final system’s efficacy by replicating bathing tasks on a mannequin in the real world. A core contribution of this work is the systematic integration of contact distributions and force data throughout the entire pipeline. Pink and green partitions delineate steps performed in the real world and simulation respectively.

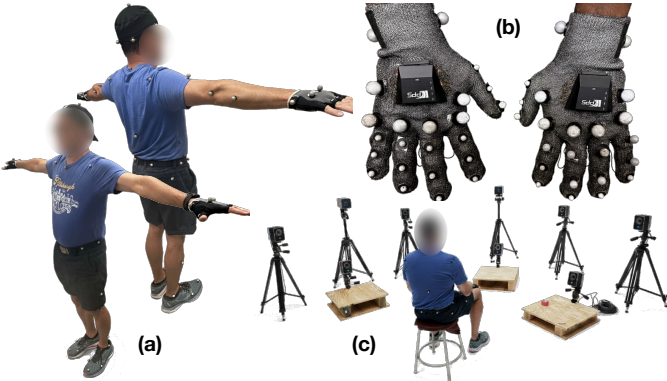


Fig. 3: Optical marker configurations used on (a) human subjects and (b) tactile force gloves. (c) Cameras are floor mounted and positioned in a semi-circle configuration.

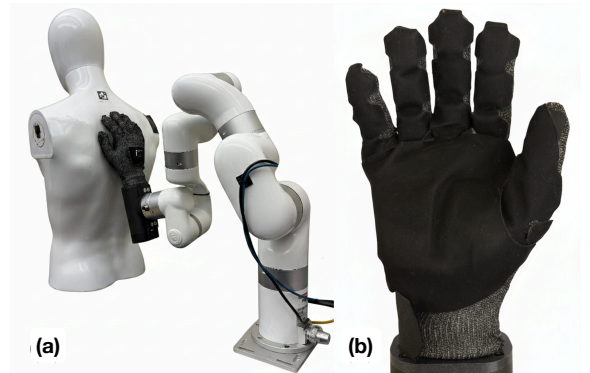


Fig. 4: (a) The robot-assisted bathing setup, consisting of a robotic arm [35], DexKit robotic hand [36], and mannequin. (b) The DexKit hand is outfitted with a tactile sensing glove [15] with grip material [37] coating the palmar side to prevent slip and moisture damage.

contact using a multi-stage pipeline. We first utilize the SMPL-X [40] and MANO [41] parametric models to estimate each subject’s body and each clinician’s hand shape parameters from their respective ROM tests using a random subset of 50 frames, as well as canonical marker position estimates relative to the baseline shape. We then estimate all pose and corrective

blend shape parameters for each bathing capture trajectory using the MoSh++ solver [42]. This process effectively converts point-based marker trajectories into complete mesh trajectories of the body and hands.

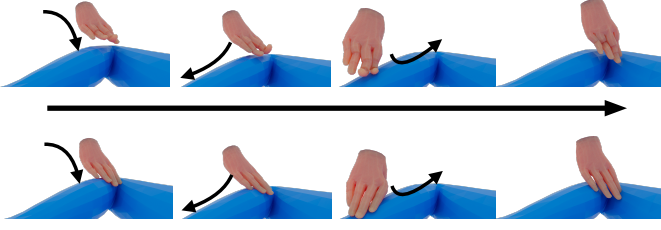


Fig. 5: Comparison film strip of a leg bathing reconstruction using (top) MoSh++ [42] and (bottom) our reconstruction pipeline. Black arrows illustrate subsequent movement directions. Our method effectively resolves finger contortion and gap artifacts without any significant parameter tuning.

However, as illustrated in Figure 5, the process routinely produces serious artifacts such as highly contorted finger poses and major gaps of separation between the hands and body, rendering it inadequate for high quality bathing reconstruction. These artifacts largely result from the necessary process of fitting bare-skin hand shape parameters to gloved captures. To mitigate these inherent discrepancies, we implement a series of straightforward, yet robust solutions.

First, to combat finger contortion artifacts, we replace the baseline fitted hand mesh with a constrained skeleton-driven reduced-DOF model that honors anatomical joint limits [43]. We use the same baseline shape parameters and drive the constrained mesh using linear blend skinning [44]. Joint angles are computed per frame using the following objective:

$$\begin{aligned} \theta^* = \arg \min_{\theta} \quad & \lambda_s \sum_{i=0}^V \|v_{bi} - v_{ci}\|_2^2 + \lambda_j \|\theta - \theta_p\|_2^2 \\ \text{s.t.} \quad & \theta_L \leq \theta \leq \theta_U \end{aligned} \quad (1)$$

where v_{bi} and v_{ci} represent corresponding vertices of the baseline and constrained hand meshes respectively, V is the total number of MANO hand vertices, θ is the constrained hand DOF vector, θ_L and θ_U are the lower and upper DOF bounds, θ_p is the existing pose, and λ_s and λ_j are weighting hyperparameters. We adopt an existing motion synthesis pipeline to solve for the full trajectory Θ^* of each constrained hand based on the per-frame pose estimates [26], and refer to this fitting process as the *shape match* pass.

We next utilize contact data to combat the gap artifacts. Starting from one-time calibrated hand taxel position estimates, we estimate corresponding body contact regions per frame using closest-point queries [45] and store the trajectories in barycentric coordinates. Next, utilizing tactile force data, we binarize the signal into activations and filter out all inactive contacts. We then perform a second *contact match* pass by adding a single additional objective term to Eq. 1:

$$\lambda_c \sum_{i=0}^C \|c_{hi} - c_{bi}\|_2^2 \quad (2)$$

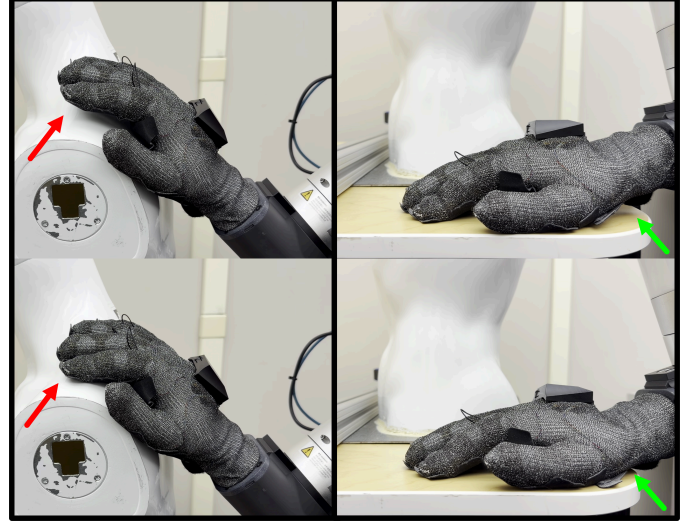


Fig. 6: Visualizations of improved surface contact following hand pose correction, comparing the original (top) and corrected (bottom) states. In the first example (left), the red arrows highlight a significant increase in finger contact. In the second (right), the green arrows indicate the hand flattening against the table as the heel moves closer to the surface.

where c_{hi} and c_{bi} represent corresponding contact points on the constrained hand and body meshes respectively, C represents the total number of taxels per hand, and λ_c is a weighting hyperparameter. Final reconstructions are complete following the contact pass. We use hyperparameters $\lambda_s = 1$, $\lambda_j = 50$ and $\lambda_c = 40$ for all captures and 100 iterations per optimization. We select these relative weights to offset dimensionality differences between V (776), C (65 / hand), and $\|\theta\|$ (29).

C. Design and Control of a Tendon-Driven Soft Hand

We select the anthropomorphic DexKit platform [36] for demonstration transfer primarily due to its compliance and secondarily for its customizability. However, because the hand is fully soft with potentially infinite DOFs and unknown kinematics, three problems must be solved to use it effectively: (1) determining a tendon routing capable of achieving desired poses, (2) determining motor positions to actuate the tendons accordingly, and (3) building an effective control space. The combination of shape, motion, and contact offered by our reconstructions helps solve each of these problems.

Starting from a dataset of human hand poses P , we first select a strictly smaller subset of candidates $S \subset P$ that sufficiently covers the full space. We do so by solving an NP-hard maximum diversity problem [46] via greedy approximation [47]. We then manually select a subset $S' \subset S$ that can be physically mimicked on DexKit under a given motor budget and commit to a tendon routing capable of generating the mimicked poses.

Next, we determine motor commands ϕ . As shown in Figure 6, calculating motor commands that adapt to contact-

induced pose corrections depends on identifying hand contact regions and accurately simulating those interactions. We can directly obtain these contact regions from the dataset.

We construct two matrices: $M_{S'}$, where each column corresponds to a pose $s \in S'$, and $M_{\phi(S')}$, where each column consists of the manually determined motor positions $\phi(s)$ required to generate that pose. For a given pose $p \in P$, we solve the following constrained least-squares QP to obtain normalized linear blend weights β^* :

$$\begin{aligned} \beta^* = \arg \min_{\beta} \quad & (M_{S'}\beta - p)^T (M_{S'}\beta - p) \\ \text{s.t.} \quad & 0 \leq \beta \leq 1 \end{aligned} \quad (3)$$

We then compute the corresponding motor positions from the optimal blend weights as $\phi = M_{\phi(S')} \beta^*$. Additional details of pose selection metrics, as well as final motor configuration and tendon routings, are available in the appendix.

D. Arm Motion Retargeting and Control

We mount the DexKit hand onto an xArm 7 [35] to perform bathing tasks on a mannequin. We deploy pre-computed motor trajectories ϕ directly to the hand, but must retarget the human hand wrist trajectories to obtain arm joint angles ψ . Unfortunately, we found that the commonly used strategy of using human wrist positions and orientations as end-effector targets [48] did not work well due to the arm's restricted workspace and the fact that hand trajectories were highly dependent on each subject's body shape and reactive motion.

Instead, we compute ψ using retargeted body contacts. Specifically, we use the TRAM video pose estimator [49] to fit SMPL-X parameters to the mannequin, and then directly roll out contact trajectories computed on the source subject. Because contact trajectories are stored in barycentric coordinates, the retargeting operation is agnostic to the target body's shape or pose parameters. We can thus always perform the rollout regardless of any shape or pose divergences between the source and target domain. Next, we replace the DexKit hand with the source MANO hand, remove all wrist DOFs, copy Θ^* from the original motion, and freeze the hand into those poses. This step allows us to re-establish correspondences between the retargeted body contacts and original hand taxels, and subsequently solve for ψ using Eq. 1. The resulting trajectory $[\{\phi_0, \psi_0\}, \{\phi_1, \psi_1\}, \dots, \{\phi_T, \psi_T\}]$ is then deployed open-loop on the real system using a modified version of the GELLO framework [50].

However, the open-loop controller frequently breaks, or alternatively applies too much, contact with the mannequin due to the gap between the real geometry and its fitted SMPL-X parameters in simulation. We create a closed-loop system by sensorizing the DexKit hand with the same tactile glove used in data capture, and utilize contact and force information from the capture to create a lower-level reactive controller. Specifically, at each time step, we apply a normalized corrective direction N_C for the end-effector along the axis intersecting the end-effector and mannequin in the real world. We convert ψ to EE positions $FK(\psi)$ rolled out at a baseline frequency

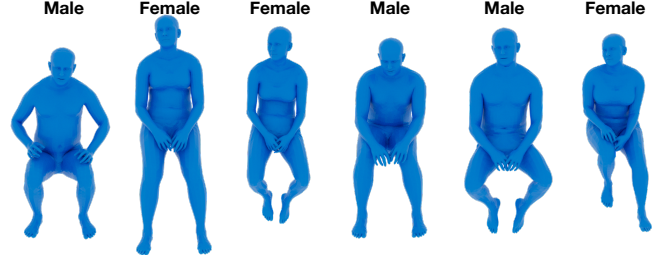


Fig. 7: MoSh++ [42] shape reconstructions of all subjects.

and apply corrective position updates at a fixed timestep Δt , creating inner loop EE position $FK(\psi) + N_C \Delta t$. Corrective movements are applied toward the body until the online sum of tactile pressures matches the original measurement, then reversed if that value is exceeded.

IV. EXPERIMENTS AND RESULTS

We discuss experiments and results concerning our three major thrusts in the proceeding subsections.

A. Dataset Statistics and Analysis

TABLE I: Distribution of all captures in our dataset segmented by body part, assistance level, and pressure intensity.

	Limited, Mild	Limited, Strong	Full, Mild	Full, Strong
Right Arm	10	8	9	7
Right Leg	6	6	3	3
Right Thigh	3	3	N/A	N/A
Right Oblique	2	2	N/A	N/A
Left Arm	7	6	7	6
Left Leg	6	6	4	3
Left Thigh	3	3	N/A	N/A
Left Oblique	2	2	N/A	N/A
Back	6	6	N/A	N/A
Neck	7	1	N/A	N/A
Face	1	N/A	N/A	N/A

Table I tabulates the distribution of captures across all subjects and clinicians in our dataset. Subjects ranged in age from 22 to 31 years, with an equal distribution of males and females. Renders of all reconstructed subject body shape parameters are available in Figure 7. Clinicians were all female, aged 30 to 32 years. The average capture duration was 2007.34 frames ($\sigma = 536.615$).

Each capture is categorized by one of two assistance variables (limited or full) and one of two pressure intensities (mild or strong). Limited assistance captures assume subjects can independently lift and hold their limbs, while full assistance captures require clinicians to manually raise and lower them throughout the bathing process. Clinicians use their discretion to differentiate between mild and strong pressure, as no explicit guidelines are provided. We also note that not all variable permutations are available per body part. For example, face bathes only consist of mild pressure demonstrations due to the region's sensitivity, areas such as the



Fig. 8: Contact coverage (yellow) over all captures of a representative (a) male and (b) female subject.

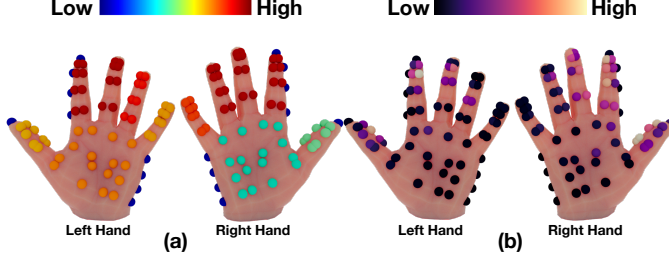


Fig. 9: Visualizations of the most commonly (a) contacted taxels and (b) applied pressures across all captures.

beck, neck, and obliques cannot possibly require full support, and multiple clinicians opted to not provide full assistance for legs because the entire limb could already be accessed from a seated position. We acknowledge that bed-bathing may provide some additional demonstrations not represented in our dataset. Finally, while we cover most parts of the body, we do not have captures of hand bathing due to tracking difficulties or of the chest and pelvic regions due to IRB restrictions. Renders of representative results are available in the supplementary video. We provide additional metrics in the proceeding subsections.

B. Visualizations of Contact Coverage

Figure 8 visualizes body coverage across all captures for two independent subjects, which we measure as all triangles

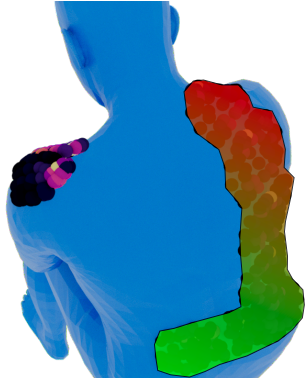


Fig. 10: Visualization of a bathing “stroke” from the (green) bottom to the (red) top of the back. The subject is stabilized by a contact distribution at the left shoulder during the stroke.

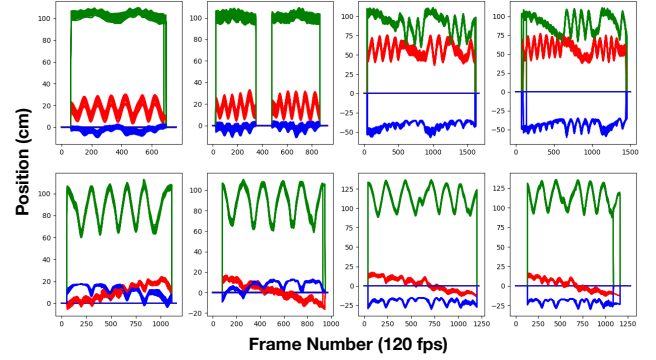


Fig. 11: Illustrations of 8 representative contact trajectories, separated into (red) X, (green) Y, and (blue) Z channels. All positions are in world coordinates. Each band is comprised of 65 individual contact point trajectories. Inactive taxels are zeroed out.

that contain a contact point at any time during a bathing trajectory. Figure 9 visualizes the most commonly contacted taxels on the left and right hands across all captures, as well as the distribution of forces across taxels. Figure 10 illustrates a sample bathing “stroke”, or surface path integral of a contact distribution over time, of the right hand up the back.

C. Quantitative Reconstruction Evaluation

TABLE II: L_2 Contact distance metric comparison between MoSh++[42] and our reconstruction.

L_2 Distance Comparisons (cm)			
	Median	Mean	St. Dev.
MoSh++	1.714	2.067	1.464
Ours	0.536	0.837	0.842

We quantitatively evaluate the quality of our reconstruction pipeline using two metrics: L_2 distances between estimated body contacts and glove taxels, and smoothness of body contact trajectories. Table II tabulates distance comparisons of our method against the original MoSh++ [42] reconstruction across 10 sample bathing demonstrations, while Figure 11 plots the trajectories of all active wiping hand body-projected taxels across multiple back bathing demonstrations. We observe high spatiotemporal correlation in the projected taxel trajectories and that oscillations coincide with multiple back-and-forth wiping passes over back segments.

D. Soft Hand Design and Control

We perform and evaluate DexKit hand design using a subset of the dataset focusing on back bathing. We collect poses and contacts from the right hand in all subject captures to build P , and perform a sweep over the number of candidates in the diversity subset S to determine the optimal sample size $|S|$. $|S|$ candidates are evaluated via MAE differentials between the original poses and their respective projections in S obtained by solving Eq. 3. We select $|S| = 50$ based on observing

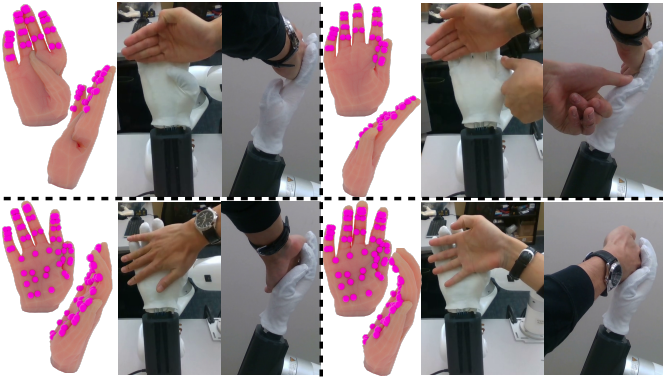


Fig. 12: Sample poses used to build $M_{\phi(S')}$. Active contacts in each pose (pink) are simulated by pressing against the hand.

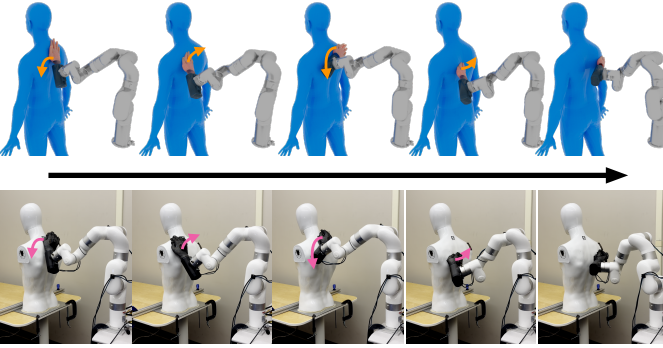


Fig. 13: Film strip of back bathing demonstration retargeted to the xArm 7 (top) in simulation using the clinician's shape-fitted hand as the end effector. The computed joint positions and pre-computed hand motor commands are (bottom) deployed on the real world setup. Orange and pink arrows illustrate subsequent movement directions.

diminishing returns in the loss curve for larger values of $|S|$, and then choose $|S'| = 20$ based on a budget of six motors. We only consider poses in contact as candidates for S and S' and store all active taxels for each selected pose. β^* weights for all trajectory poses are computed by solving Eq. 3 using the resulting $M_{S'}$.

We next build $M_{\phi(S')}$ and compute ϕ . We manually determine motor positions for each candidate pose without contact, and then adjust positions while pressing the hand at the regions of contact designated by the active taxels. Figure 12 illustrates several poses used during calibration.

E. Contact Retargeting and Real World Deployment

Figure 13 illustrates the xArm positions computed following a retargeted contact trajectory rollout, as well as deployment of that trajectory on our real world platform. ψ and ϕ commands are deployed together in the control loop, and the original human demonstration is slowed from 120 to 6 Hz (20X). Complete results are available in the supplementary video. We provide additional Sim2Real alignment details in the appendix.

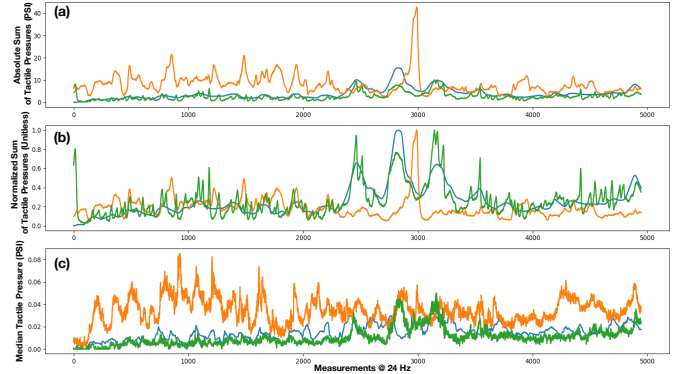


Fig. 14: Overlaid plots of the (blue) original human, (orange) robot open loop, and (green) robot closed loop tactile glove pressures during a back bathing task. We provide (a) absolute sums, (b) normalized sums, and (c) median intensities calculated across the full taxel array per time step. Note that the spike in pressure at the start of closed loop data collection is due to initialization noise rather than a consequence of trajectory execution.

Figure 14 compares the recorded glove pressure signals across open-loop rollout, closed-loop rollout, and the original human demonstration. As expected, the integrated tactile pressure for the open-loop rollouts is significantly higher than both the human demonstration and closed-loop results for the majority of the trajectory. The unsafe operating conditions indicated by these elevated pressure levels are further evidenced by the mannequin vibrations visible in the supplementary video. Closed-loop control substantially curtails such issues, as quantitatively evidenced in the plots and qualitatively by minimal mannequin vibrations in the video. Closed-loop pressures also reasonably track the original human demonstration, particularly when viewed on a normalized scale.

However, the closed-loop controller deviates from the human demonstration in absolute pressure and especially spatial distribution, as indicated by the comparison of median values. The distributional misalignment is an inherent trade-off of the controller design, as optimizing for an aggregate pressure sum does not necessitate parity in the underlying taxel-wise distribution. Furthermore, the attenuated dynamic range is a byproduct of necessary safety constraints in real world deployment. Specifically, we found that targeting a strict absolute pressure induces significant arm oscillations and mitigated this behavior by implementing tolerance bounds around the target instead. We also identified a failure mode where friction-induced finger curling during wiping motions could lead to mechanical damage. If curling occurs while the controller commands an inward corrective force, the system may drive the arm further into the mannequin in a futile attempt to reach a pressure target that the compromised finger geometry cannot achieve. To mitigate this, we implemented a temporal safety timeout for unreached pressure setpoints. Interestingly, this failure mode was absent in open-loop trials; because the controller advances through the trajectory regardless of

feedback, the hand’s passive compliance allowed fingers to recover from compression during lateral movements.

V. DISCUSSION

Several key insights emerged during the reconstruction and transfer process. We outline our primary findings through the dataset, lessons learned, and current limitations below.

A. Data Insights

As strongly evidenced by video results and Figures 10 and 11, bathing contact trajectories are *highly* correlated, smooth, and dynamic. Modeling contacts as individual points, instantaneous static entities, or simple constraints, as is still commonly done by many long-standing analysis techniques [51, 52, 53] and modern simulators [54, 55], is highly limiting for representing our captured data. Contact models based on surface area [21, 56] or volume [57]—when paired with well-defined velocity parameters—offer significantly more promise for capturing the sustained interactions necessary for bathing.

The recorded tactile signals for mild pressure demonstrations, regardless of clinician or body part, are quite low (in multiple cases below the minimum sensitivity of 0.04 N) and thus largely indistinguishable from noise. We therefore strongly advise prioritizing high contact sensitivity in hardware design for low-pressure data acquisition, or alternatively focusing efforts on strong pressure captures.

Unsurprisingly, we found that signals between adjacent taxels tend to be highly correlated regardless of the task. We argue there is little value gained from high resolution sensing of regions unaffected by pose changes (e.g. the proximal phalanx of an individual finger). In contrast, hand shapes induced by pose changes, as well as hand shape in relation to local body shape, have a much more significant impact on pressure distributions. Therefore, we argue that thinking of sensor coverage in terms of *regions between joints* is more valuable for bathing than intra-region resolution.

Interestingly, in captures where both hands were utilized, the majority of high pressure measurements came from the supporting hand rather than the wiping hand. This observation indicates that body stabilization requires substantially more force than wiping, and therefore suggests that asymmetry may be a desirable characteristic of bimanual bathing systems. For example, the stabilizing arm and hand can utilize high torque, low gear-reduced actuators and sport limited DOFs, while the wiping hand can trade off power for speed.

Finally, we observed a *substantial* amount of patient movement, both voluntary and in response to bathing forces. While our initial proof-of-concept transfers utilized a static mannequin, any system intended for human application *must* account for constant body re-positioning.

B. Lessons Learned

While tactile information does help determine precise times and magnitude of contact, its utility is limited without context from motion and shape. Tactile data alone cannot determine

between which bodies contact is happening or *why* forces change over time. Both issues made identifying true contact events between the hand and body difficult to automate. Multiple contact events between the thumb and thenar region of the palm, for example, resulted from gripping the wipe rather than applying body pressure. Drops in pressure could be attributed to either breaking contact or simply reducing contact force. Low, but increasing pressure signals could be actual contact events or drift noise. We found that while motion and shape data are vital for clarifying such events, they are not a total solution; a holistic analysis of all three channels together is still required to accurately interpret results.

We also found that compliance was *of paramount importance* in arm-mounted bathing systems. Sustained, dense-contact interactions permit little room for trajectory mismatch errors, which are especially hard to avoid during sim2real rollouts. Although the Dexkit hand’s material compliance mitigated these issues during our experiments, we contend that incorporating compliance into arm control is essential for safe and effective interaction with human subjects.

Finally, we ultimately found that a flexor-only finger tendon configuration was insufficient to capture all necessary hand motions during rollouts. Passive compliance did not always allow the hand to escape friction-induced finger curling during wiping motions. Including dedicated extensor tendons would help mitigate such issues, but would be difficult to do under limited motor budgets.

C. Limitations and Future Work

Our system currently contains major current limitations. Our closed-loop strategy, for example, is highly limited by its use of tactile pressure sums rather than the distribution over individual taxels. The primary barrier to distributed force control is the significant discrepancy between human and Dexkit hand responses when subjected to similar pressure stimuli. While existing methods mitigate this issue through careful placement of sensors, selection of sensing modality, and calibration to align readings between similar human and robot trajectories [18, 58], such approaches tend to be brittle and require significant engineering overhead. There remains an opportunity to more directly model the inherent discrepancies in hand materials. We propose that *retargeting force distributions* to different geometries and material properties is a promising direction of future research.

Our approach’s simulation-to-reality gap is substantial. While our force data is grounded purely in the real world, estimates of the xArm configurations and hand controller blend weights require substantial calibration due to mismatches in the mannequin vs. SMPL-X fitted model and MANO-reconstructed hand vs. Dexkit hand control spaces respectively. We also do not account for compliance in the Dexkit hand, reconstructed human body, or reconstructed human hand, which together can complicate accurately reproducing clinician-induced sensations beyond taxel observations.

Our data does not differentiate between normal and shear forces. This limitation can be attributed to our glove’s ca-

capacitive sensing modality; alternatives such as visual [59] and magnetic [18] sensing can provide such measurements, albeit at the cost of form factor, lower SNR, reliability, or interpretability. While force direction will not impact our reconstruction pipeline, the information can be useful in online control scenarios such as helping more carefully manage applied frictional forces during sliding or escaping the identified curling failure mode.

Because of these and other limitations, our system is not yet ready for deployment on human subjects. Even with mechanical compliance, the primary bottleneck is perception. Specifically, no existing method achieves fully online human pose and shape estimation while maintaining the sub-frame latency required for live interaction, and our current assumption of a static body will not extend to a human subject. We argue that this component is vital both for retargeting contact trajectories from existing demonstrations and for enabling the system to react to human pose changes. Addressing this challenge could pave the way for the first whole-body bathing system capable of learning directly from human demonstrations.

VI. CONCLUSION

We have presented a straightforward, but effective framework for high fidelity capture and reconstruction of human bathing demonstrations, as well as methods to facilitate transfer of demonstrations across multiple levels of the robotics stack. Leveraging our framework, we have created the inaugural dataset of whole-body bathing demonstrations performed by trained clinicians on human subjects. Importantly, we have shown that utilizing contact regions as a key processing primitive effectively addresses many pain points in the reconstruction and transfer pipelines. We hope that the contributions of our methods, dataset, and code will spur further research in pHRI and ultimately make progress towards bringing useful and reliable robots into high-value clinical settings.

ACKNOWLEDGMENTS

We thank Justin C. Macey for assistance with data capture and cleanup, as well as Giorgio Becherini, Shashank Tripathi, and Alpár Cseke for help in setting up the MoSh++ solver. Research reported in this publication was partially supported by the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under award number R01EB036842 and an award from the RAI Institute.

REFERENCES

- [1] D. D. Dunlop, S. L. Hughes, and L. M. Manheim. Disability in activities of daily living: patterns of change and a hierarchy of disability. *American journal of public health*, 87(3):378–383, 1997.
- [2] S. Katz, A. B. Ford, R. W. Moskowitz, B. A. Jackson, and M. W. Jaffe. Studies of illness in the aged: the index of adl: a standardized measure of biological and psychosocial function. *Journal of the American Medical Association*, 185(12):914–919, 1963.
- [3] T. M. Gill, A. Guo, and H. G. Allore. The epidemiology of bathing disability in older persons. *Journal of the American Geriatrics Society*, 54(10):1524–1530, 2006.
- [4] L. R. Czuba, C. M. Sommerich, and S. A. Lavender. Ergonomic and safety risk factors in home health care: Exploration and assessment of alternative interventions. *Work*, 42(3):341–353, 2012.
- [5] A. R. Darragh, C. M. Sommerich, S. A. Lavender, K. J. Tanner, K. Vogel, and M. Campo. Musculoskeletal discomfort, physical demand, and caregiving activities in informal caregivers. *Journal of Applied Gerontology*, 34(6):734–760, 2015.
- [6] Z. Erickson, H. M. Clever, V. Gangaram, G. Turk, C. K. Liu, and C. C. Kemp. Multidimensional capacitive sensing for robot-assisted dressing and bathing. In *IEEE International Conference on Rehabilitation Robotics*, pages 224–231, 2019.
- [7] F. Liu, V. Patil, Z. Erickson, and Z. Temel. Characterization of a meso-scale wearable robot for bathing assistance. In *IEEE International Conference on Robotics and Biomimetics*, pages 2146–2152. IEEE, 2022.
- [8] R. Madan, S. Valdez, D. Kim, S. Fang, L. Zhong, D. T. Virtue, and T. Bhattacharjee. Rabbit: A robot-assisted bed bathing system with multimodal perception and integrated compliance. In *Proceedings of the 2024 ACM/IEEE international conference on human-robot interaction*, pages 472–481, 2024.
- [9] I. Huang, D. Chow, and R. Bajcsy. Soft tactile contour following for robot-assisted wiping and bathing. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 7797–7802, 2022.
- [10] A. C. Dometios, X. S. Papageorgiou, A. Arvanitakis, C. S. Tzafestas, and P. Maragos. Real-time end-effector motion behavior planning approach using on-line point-cloud data towards a user adaptive assistive bath robot. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5031–5036, 2017.
- [11] A. Zlatintsi, I. Rodomagoulakis, P. Koutras, A. C. Dometios, V. Pitsikalis, C. S. Tzafestas, and P. Maragos. Multimodal signal processing and learning aspects of human-robot interaction for an assistive bathing robot. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3171–3175, 2018.
- [12] C. H. King, T. L. Chen, A. Jain, and C. C. Kemp. Towards an assistive robot that autonomously performs bed baths for patient hygiene. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 319–324, 2010.
- [13] S. Sundaram, P. Kellnhofer, Y. Li, J. Y. Zhu, A. Torralba, and W. Matusik. Learning the signatures of the human grasp using a scalable tactile glove. *Nature*, pages 698–702, 2019.
- [14] Y. Luo, C. Liu, Y. J. Lee, J. DelPreto, K. Wu, M. Foshey, D. Rus, T. Palacios, Y. Li, a. Torralba, et al. Adaptive tactile interaction transfer via digitally embroidered smart gloves. *Nature Communications*, 15(1):868, 2024.

- [15] Pressure Profile Systems. TactileGlove: Hand pressure and force measurement. <https://pressureprofile.com/body-pressure-mapping/tactile-glove>, 2026. Accessed: 2026-01-26.
- [16] H. S. Fang, B. Romero, Y. Xie, A. Hu, B. R. Huang, J. Alvarez, M. Kim, G. Margolis, K. Anbarasu, M. Tomizuka, et al. Dexop: A device for robotic transfer of dexterous human manipulation. *arXiv preprint arXiv:2509.04441*, 2025.
- [17] A. Adeniji, Z. Chen, V. Liu, V. Pattabiraman, R. Bhirangi, S. Haldar, P. Abbeel, and L. Pinto. Feel the force: Contact-driven learning from humans. *arXiv preprint arXiv:2506.01944*, 2025.
- [18] J. Yin, H. Qi, Y. Wi, S. Kundu, M. Lambeta, W. Yang, C. Wang, T. Wu, J. Malik, and T. Hellebrekers. Osmo: Open-source tactile glove for human-to-robot skill transfer. *arXiv preprint arXiv:2512.08920*, 2025.
- [19] Y. Mao, U. Yoo, Y. Yao, S. N. Syed, L. Bondi, J. Francis, J. Oh, and J. Ichnowski. Visuo-acoustic hand pose and contact estimation. *arXiv preprint arXiv:2508.00852*, 2025.
- [20] C. Xing, H. Li, Y. L. Wei, T. A. Ren, T. Tu, Y. Lin, E. Schumann, W. S. Zheng, and M. R. Cutkosky. Taccap: A wearable fbg-based tactile sensor for seamless human-to-robot skill transfer. *arXiv preprint arXiv:2503.01789*, 2025.
- [21] A. S. Lakshmipathy, N. Feng, Y. X. Lee, M. Mahler, and N. S. Pollard. Contact edit: Artist tools for intuitive modeling of hand-object interactions. *ACM Transactions on Graphics*, 42(4), 2023.
- [22] D. Turpin, L. Wang, E. Heiden, Y. C. Chen, M. Macklin, S. Tsogkas, S. Dickinson, and A. Garg. Grasp’d: Differentiable contact-rich grasp synthesis for multi-fingered hands. In *European Conference on Computer Vision*, pages 201–221. Springer, 2022.
- [23] S. Brahmbhatt, A. Handa, J. Hays, and D. Fox. Contactgrasp: Functional multi-finger grasp synthesis from contact. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2386–2393. IEEE, 2019.
- [24] S. Christen, M. Kocabas, E. Aksan, J. Hwangbo, J. Song, and O. Hilliges. D-grasp: Physically plausible dynamic grasp synthesis for hand-object interactions. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20577–20586, 2022.
- [25] Y. Wu, J. Wang, Y. Zhang, S. Zhang, O. Hilliges, F. Yu, and S. Tang. Saga: Stochastic whole-body grasping with contact. In *European Conference on Computer Vision*, pages 257–274. Springer, 2022.
- [26] A. S. Lakshmipathy, J. K. Hodgins, and N. S. Pollard. Kinematic motion retargeting for contact-rich anthropomorphic manipulations. *ACM Transactions on Graphics*, 44(2):1–20, 2025.
- [27] Z. Mandi, Y. Hou, D. Fox, Y. Narang, A. Mandlekar, and S. Song. Dexmachina: Functional retargeting for bimanual dexterous manipulation. *arXiv preprint arXiv:2505.24853*, 2025.
- [28] C. Pan, C. Wang, H. Qi, Z. Liu, H. Bharadhwaj, A. Sharma, T. Wu, G. Shi, J. Malik, and F. Hogan. Spider: Scalable physics-informed dexterous retargeting. *arXiv preprint arXiv:2511.09484*, 2025.
- [29] T. Pang, H. J. T. Suh, L. Yang, and R. Tedrake. Global planning for contact-rich manipulation via local smoothing of quasi-dynamic contact models. *IEEE Transactions on robotics*, 39(6):4691–4711, 2023.
- [30] X. Cheng, E. Huang, Y. Hou, and M. T. Mason. Contact mode guided motion planning for quasidynamic dexterous manipulation in 3d. In *International Conference on Robotics and Automation*, pages 2730–2736. IEEE, 2022.
- [31] 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*, pages 581–600. Springer, 2020.
- [32] Z. Fan, O. Taheri, D. Tzionas, M. Kocabas, M. Kaufmann, M. J. Black, and O. Hilliges. Arctic: A dataset for dexterous bimanual hand-object manipulation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12943–12954, 2023.
- [33] J. DelPreto, C. Liu, Y. Luo, M. Foshey, Y. Li, A. Torralba, W. Matusik, and D. Rus. Actionsense: A multimodal dataset and recording framework for human activities using wearable sensors in a kitchen environment. *Advances in Neural Information Processing Systems*, 35:13800–13813, 2022.
- [34] Y. R. Song, J. Li, R. Fu, D. Murphy, K. Zhou, R. Shiv, Y. Li, H. Xiong, C. E. Owens, Y. Du, et al. Opentouch: Bringing full-hand touch to real-world interaction. *arXiv preprint arXiv:2512.16842*, 2025.
- [35] UFACTORY. UFACTORY xArm 7 Robotic Arm. <https://www.ufactory.us/product/ufactory-xarm-7>, 2026. Accessed: 2026-01-26.
- [36] J. P. King, H. Ahluwalia, I. B. Heredia-Marin, K. McClary, E. Zuo, M. Baratz, J. E. Colgate, J. Clark, M. Orta Martinez, and N. S. Pollard. Dexkit: A hardware platform for multi-finger dexterous robotics. In *Proceedings of the 19th International Symposium on Experimental Robotics*, 2025.
- [37] 3M™. Gripping Material TB641. https://www.3m.com/3M/en_US/p/d/b40069682/, 2026. Accessed: 2026-01-26.
- [38] L. K. Baxter. Capacitive sensors. *IEEE Design and Applications*, 1997.
- [39] Vicon Motion Systems. V16 camera specifications - Vantage documentation. <https://help.vicon.com/space/Vantage/15041618/V16+camera+specifications>, 2026. Accessed: 2026-01-26.
- [40] G. Pavlakos, V. Choutas, N. Ghorbani, T. Bolkart, A. A. A. Osman, D. Tzionas, and M. J. Black. Expressive body capture: 3d hands, face, and body from a single image. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10975–10985, 2019.
- [41] J. Romero, D. Tzionas, and M. J. Black. Embodied

- hands: Modeling and capturing hands and bodies together. *ACM Transactions on Graphics*, 36(6), 2017.
- [42] N. Mahmood, N. Ghorbani, N. F. Troje, G. Pons-Moll, and M. J. Black. Amass: Archive of motion capture as surface shapes. In *IEEE International Conference on Computer Vision*, pages 5442–5451, 2019.
- [43] D. J. Magee. *Orthopedic Physical Assessment*. Elsevier Health Sciences, 2013.
- [44] N. Magnenat-Thalmann, R. Laperrière, and D. Thalmann. Joint-dependent local deformations for hand animation and object grasping. In *Graphics Interface*, pages 26–33, 1989.
- [45] R. Sawhney. Fcpw: Fastest closest points in the west, 2021.
- [46] C. C. Kuo, R. Glover, and K. S. Dhir. Analyzing and modeling the maximum diversity problem by zero-one programming. *Decision Sciences*, 24(6):1171–1185, 1993.
- [47] J. B. Ghosh. Computational aspects of the maximum diversity problem. *Operations Research Letters*, 19(4):175–181, 1996.
- [48] A. Sivakumar, K. Shaw, and D. Pathak. Robotic telekinesis: Learning a robotic hand imitator by watching humans on youtube. *arXiv preprint arXiv:2202.10448*, 2022.
- [49] Y. Wang, Z. Wang, L. Liu, and K. Daniilidis. Tram: Global trajectory and motion of 3d humans from in-the-wild videos. In *European Conference on Computer Vision*, pages 467–487, 2024.
- [50] P. Wu, Y. Shentu, Z. Yi, X. Lin, and P. Abbeel. Gello: A general, low-cost, and intuitive teleoperation framework for robot manipulators. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 12156–12163, 2024.
- [51] M. T. Mason. *Mechanics of robotic manipulation*. MIT press, 2001.
- [52] C. Ferrari, J. Canny, et al. Planning optimal grasps. In *Proceedings., 1992 IEEE International Conference on Robotics and Automation, 1992.*, volume 3, pages 2290–2295. IEEE, 1992.
- [53] K. M. Lynch and F. C. Park. *Modern robotics*. Cambridge University Press, 2017.
- [54] E. Todorov, T. Erez, and Y. Tassa. Mujoco: A physics engine for model-based control. In *IEEE/RSJ International conference on intelligent robots and systems*, pages 5026–5033, 2012.
- [55] V. Makoviychuk, L. Wawrzyniak, Y. Guo, M. Lu, K. Storey, M. Macklin, D. Hoeller, N. Rudin, A. Allshire, A. Handa, et al. Isaac gym: High performance gpu-based physics simulation for robot learning. *arXiv preprint arXiv:2108.10470*, 2021.
- [56] A. Lakshminpathy, D. Bauer, C. Bauer, and N. S. Pollard. Contact transfer: A direct, user-driven method for human to robot transfer of grasps and manipulations. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 6195–6201. IEEE, 2022.
- [57] R. Elandt, E. Drumwright, M. Sherman, and A. Ruina. A pressure field model for fast, robust approximation of net contact force and moment between nominally rigid objects. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 8238–8245. IEEE, 2019.
- [58] Y. Wi, J. Yin, E. Xiang, A. Sharma, J. Malik, M. Mukadam, N. Fazeli, and T. Hellebrekers. Tactalign: Human-to-robot policy transfer via tactile alignment. *arXiv:2602.13579*, 2026.
- [59] W. Yuan, S. Dong, and E. H. Adelson. Gelsight: High-resolution robot tactile sensors for estimating geometry and force. *Sensors*, 17(12):2762, 2017.

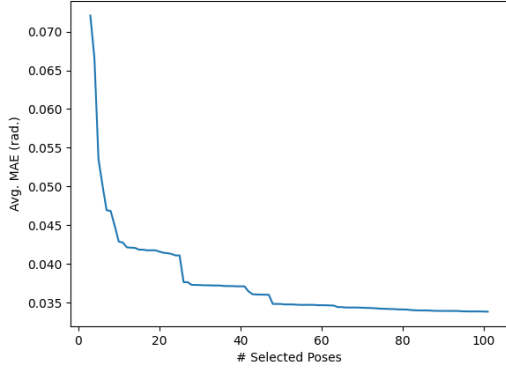


Fig. 15: Complete MAE loss curve over the parameter sweep of $|S|$.

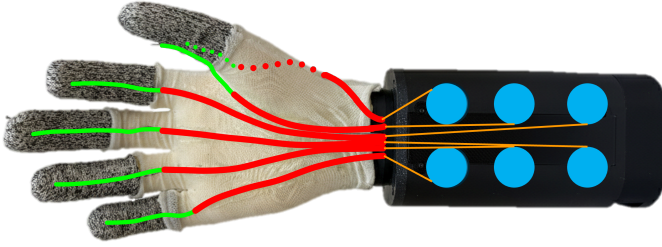


Fig. 16: The DexKit hand is comprised of a cast foam hand with a textile “skin”. Tendons are sewn into the skin (green), and routed along the palm through captive PTFE sheaths (red). The hand shown has six tendons, five flexors on the fingers and thumb, as well as an adductor on the thumb which is routed along the back. The hand interfaces with six servo motors packaged in a housing block at the wrist (black). Tendons in the block (orange) route to pulleys (blue) on the motors which reel in the tendons to actuate the hand.

APPENDIX

We utilize this supplemental document to detail our real world setups, constraints, and design choices concerning the DexKit and xArm platforms.

A. DexKit Pose Selection Metrics

Figure 15 illustrates the complete loss curve over a sweep of $|S| = 1$ to $|S| = 100$. The size of all right hand back bathing poses $|P| = 9491$. While there are further dips in the curve at values $|S| > 100$, and it is possible to achieve zero error by choosing $|S| = |P|$, we found that many poses were qualitatively repetitive. Errors of re-projection MAE losses computed using the finally constructed $M_{S'}$ were normally distributed with parameters $(\mu, \sigma) = (0.04477, 0.01928)$. Although these statistics suggest effective computed blend weights β^* , we note that metrics concerning MANO hand poses are not necessarily strong measures for evaluating DexKit pose coverage.

B. Dexkit Motor Configuration and Tendon Routing

Figure 16 provides schematics of the final Dexkit motor and tendon configuration. We select a flexor-only configuration for

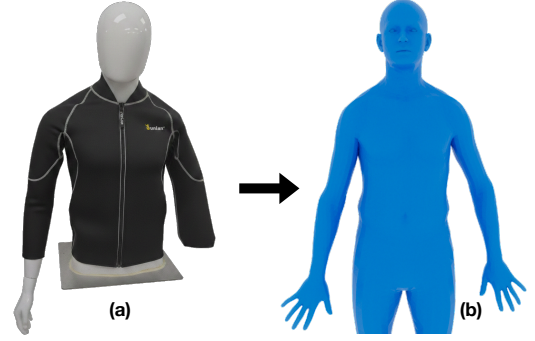


Fig. 17: Illustrations of the (a) mannequin input image and (b) extracted SMPL-X proxy mesh from TRAM [49].

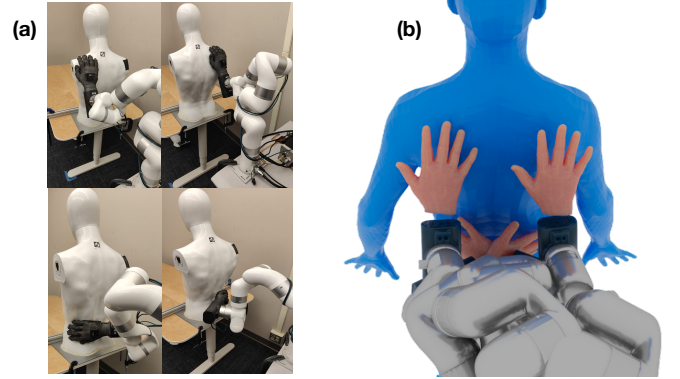


Fig. 18: (a) xArm calibration poses collected in the real world are (b) superimposed for mannequin positioning in simulation.

the index, middle, ring, and pinky fingers, and an adduction routing for the thumb to pull in closer to the palm thenar region. We found the adduction routing more useful than a flexure routing because the flexed thumb interfered with the palm making contact with the mannequin during wiping the relatively flat back regions. The total system mass is roughly 0.25 kg, which is well within the xArm payload limit.

C. Sim2Real Alignment Details

Figure 17 depicts the mannequin image and TRAM-reconstructed SMPL-X body shape, averaged across 10 frames of video. Notably, we found that attaching at least one arm and placing clothing on the mannequin was necessary because the model otherwise failed to identify the structure as a human. We only utilize the baseline body shape parameters, ignoring blend shape corrections and root pose estimates.

Instead, we estimate the root position through real-world calibration. We manually adjust the xArm into four distinct poses to estimate the “boundaries” of the mannequins back, which include the left and right shoulders as well as the left and right hip segments. We then manually align the mannequin proxy in simulation relative to the corresponding xArm positions, and note that we intentionally offset the proxy mesh slightly further from the hands. This accounts for shape discrepancies and ensures the arm biases toward overreaching rather than falling short. Figure 18 illustrates the process.