Motion Retargeting for Grasping Tasks

Mapping synergies from humans to robotic hands with dissimilar kinematics

A.A. 2024 - 2025

Matteo Zamponi
Luca Colamarino
Giuseppe D'Addario
Federico Tranzocchi
Sapienza University of Rome

Contents

Chapter 1

Introduction

Robotic hands are becoming increasingly common in research and industry. Modern robotic hands often have many degrees of freedom (DoF) to imitate the dexterity of the human hand, which has over 20 DoF. However, controlling these high-DoF robotic hands is a complex task: while the human brain controls the hand easily using coordinated patterns called *synergies* [1], explicitly controlling 15 or more motors of a robotic hand would be impractical.

This project addresses the problem of *motion retargeting* from a human hand to a non-anthropomorphic robotic hand. The goal is to leverage human hand synergies to control a robotic hand with a different kinematic structure to that of the human hand. To do this, we use an object-based mapping approach called the *virtual sphere* method [2], which focuses on the interaction between the hand and the object being manipulated.

1.1 The correspondence problem

A major challenge in teleoperation is the *correspondence problem*. Human hands and robotic hands are rarely identical: they usually have different bone lengths, different types of joints, and a different number of fingers. Because of these differences, standard mapping techniques like mapping joint angles directly or mapping fingertip positions fail to accurately translate human motion.

To solve this, we adopt an approach proposed by Gioioso et al. [2] that works in the *object domain*. Instead of mapping the anatomy of the hand directly, we map the effect that the hand has on a virtual object (a sphere) held in the hand.

1.2 Synergy-driven input

To control the robotic hand, we first need a reliable reconstruction of the human hand pose. In this project, we use the Weart TouchDIVER G1 haptic glove in Figure 1.1, which provides the closure of three fingers (thumb, index, middle) and the abduction of the thumb, plus other measurements that are not of interest for this work.

To reconstruct the full hand pose from this limited input, we rely on a reconstruction system developed in a previous work by Primiceri et al. [4]. Their system uses neural networks to estimate the full human hand pose from the sparse glove data, based on the theory of *postural synergies* [1]. We take this reconstructed hand pose as the input for our motion retargeting system, focusing on the mathematical translation of this motion to the robotic hand.

Figure 1.1: The Weart TouchDIVER G1 haptic glove. Image source: [3].

1.3 Objective

Our main goal is to implement and validate a robust pipeline that allows a human operator to control robotic hands with different kinematics in real-time. Specifically, we present an algorithm that creates a virtual sphere inside the human hand and maps its deformation and movement to the robotic hand. Thanks to the mapping in the domain of the manipulated object, we create a generalized kinematic solver that can handle different robotic structures (e.g., from a 5-fingered hand to a 3-fingered one, or a 2-fingered gripper) without needing to redesign the mapping for each case. Finally, we evaluate the performance of our system by *TODO: add evaluation details*.

1.4 Structure

The rest of this report is structured as follows. In Chapter 2 we review the relevant literature on hand synergies and motion retargeting techniques. In Chapter 3 we provide a mathematical formulation of the virtual sphere method and our implementation details. In Chapter 4 we describe the system architecture, including the hardware and software components used. In Chapter 5 we present the results of our experiments and evaluate the performance of the retargeting system. Finally, in Chapter 6 we summarize our findings and discuss potential future work.

Chapter 2

Background

In this chapter, we review the theoretical foundations and existing methodologies that form the basis of our work. We first discuss the biomechanical principles of human hand control, specifically the concept of postural synergies. We then analyze standard approaches for mapping human motion to robotic hands, highlighting their limitations when dealing with dissimilar kinematics. Finally, we introduce the object-based retargeting approach, which provides the theoretical framework for the virtual sphere method used in this project.

2.1 Postural synergies

The human hand is a kinematic structure with more than 20 degrees of freedom (DoF) controlled by a complex network of muscles and tendons. Despite this mechanical complexity, humans are able to grasp objects of different shapes and sizes with ease and dexterity. Neuroscientific studies suggest that the central nervous system simplifies the control of the hand by coordinating the movement of multiple joints through a reduced set of control variables known as *postural synergies* [1]. These synergies represent patterns of joint coordination that capture the most significant variations in hand posture during grasping tasks.

In their study, Santello et al. [1] analyzed the hand postures of different subjects while grasping imaginary objects. By applying *Principal Component Analysis* (PCA) to the collected joint angle data, they identified that a large portion of the variance in hand postures could be explained by a very small number of principal components. Specifically:

- the first principal components (*synergies*) account for more than 80% of the variance in hand posture;
- the first synergy corresponds to the coordinated flexion and extension of all fingers (opening and closing the hand), resembling a power grasp *TODO: add figure*;
- the second synergy accounts for the abduction of the fingers and the opposition of the thumb, effectively controlling the arching of the palm *TODO: add figure*;
- the remaining synergies can be used to fine-tune the hand posture for specific grasp types, but contribute increasingly less to the overall variance.

This finding is fundamental to our work: it implies that we do not need to measure every single joint of the human hand to understand its pose. Instead, we can capture the underlying correlations to reconstruct the full hand posture from the sparse data provided by the Weart glove, as demonstrated by the neural network approach [4] that we use.

2.2 Standard motion retargeting strategies

Retargeting human hand motion to a robotic hand is a classic problem in robotics, often referred to as the *correspondence problem*. The challenge arises from the fact that robotic hands typically have different kinematic structures, sizes, and joint limits compared to the human hand. In literature, we can find two main strategies to address this problem.

Joint-to-joint mapping This is the most direct approach, where the joint angles of the human hand are mapped one-to-one to the corresponding joints of the robotic hand. If the robot is anthropomorphic, we can map human joints $\boldsymbol{q}_h$ directly to robot joints $\boldsymbol{q}_r$ using a linear transformation; if the robot has fewer joints, we need a mapping function to approximate the motion.

This method is straightforward and relatively easy to implement, but it fails with non-anthropomorphic hands (e.g., a 3-fingered gripper). Since the kinematic chains differ, applying human angles directly can lead to unnatural or infeasible robot postures, self-collisions, or the inability to grasp objects properly.

Fingertip (Cartesian) mapping To overcome the limitations of joint-to-joint mapping, another common approach is to focus on the positions of the fingertips rather than the joint angles. The Cartesian coordinates of the human fingertips $\boldsymbol{p}_h$ are computed through forward kinematics and then used as target positions for the robot's fingertips $\boldsymbol{p}_r$. An inverse kinematics solver is then employed to find the joint angles $\boldsymbol{q}_r$ that achieve these positions.

While this ensures that the fingertips reach the target, it completely ignores the internal configuration of the hand, potentially leading to unnatural grasps or excessive joint movements. Additionally, if the robot has fewer fingers than the human hand, it becomes unclear how to map multiple human fingertips to fewer robotic ones: a fingertip position reachable by a human might be a singular or unreachable configuration for the robot.

2.3 Object-based retargeting

To address the limitations of the standard retargeting methods, Gioioso et al. [2] proposed an approach defined in the *object domain*: instead of mapping the hand itself, either in joint or Cartesian space, we map the effect the hand has on the object being manipulated. The key idea is that the primary goal of the hand is to interact with objects, so by focusing on the object, we can achieve more natural and effective grasps.

Since the object might not physically exist during teleoperation (e.g., in virtual reality scenarios), a *virtual object* (a sphere) is introduced in the grasping process. The method proceeds in the following steps:

1. A virtual sphere is mathematically fitted inside the human hand, defined by a set of reference points (e.g., fingertips and palm center). As the human hand moves via synergies, this sphere *translates*, *rotates*, and *deforms* accordingly.

2. The motion of the human sphere is scaled to match the size of the robotic hand, allowing a large human hand to control a small robotic gripper, or vice versa.

3. The robot is commanded to move its joints such that its own virtual sphere mimics the transformation of the human's sphere.

This object-based retargeting method is *independent* of the kinematic structure of both hands. It captures the intention of the grasp, namely whether the user is squeezing (shrinking the sphere) or moving the hand (translating the sphere), and it applies to the robot regardless of its number of fingers or joint structure.

Chapter 3

Framework

In this chapter, we present the mathematical formulation of the proposed motion retargeting framework. We detail the definition of the virtual sphere, the construction of the interaction matrix that relates hand motion to object deformation, and the derivation of the control law used to drive the robotic hand. Finally, we describe the redundancy resolution strategy employed to optimize the configuration of the robotic hand during manipulation task.

3.1 Reference points and sphere definition

The core concept of the object-based mapping is to abstract the hand's motion into the motion of a virtual object. We model this object as a *virtual sphere* defined by a set of reference points on the hand.

Let $\boldsymbol{p}_h \in \mathbb{R}^{3N_h}$ be the vector of the Cartesian positions of N_h reference points on the human hand. Similarly, let $\boldsymbol{p}_r \in \mathbb{R}^{3N_r}$ be the vector of N_r reference points on the robotic hand. At any time step t, the virtual sphere is defined as the *minimum enclosing ball*, i.e., the sphere of smallest radius that contains all reference points. Such a sphere is characterized by its center $\boldsymbol{o} \in \mathbb{R}^3$ and radius $r \in \mathbb{R}$.

The objective of the retargeting algorithm is to impose that the virtual sphere of the robotic hand mimics the rigid-body motion (translation and rotation) and the non-rigid deformation (scaling) of the human virtual sphere. To account for the size difference between the two hands, we introduce a scaling factor k_{sc}:

$$k_{sc} = \frac{r_r}{r_h} \tag{3.1}$$

where r_r and r_h are the radii of the robotic and human virtual spheres in their initial reference configurations, respectively.

3.2 Interaction matrix

To map the motion of the reference points to the motion of the sphere, we use the definition provided by Gioioso et al. [2]. The velocity of a generic reference points $\boldsymbol{p}_i$ can be expressed as a function of the sphere's linear velocity $\dot{\boldsymbol{o}}$, angular velocity $\boldsymbol{\omega}$, and rate of change of radius $\dot{r}$:

$$\dot{\boldsymbol{p}}_i = \dot{\boldsymbol{o}} + \boldsymbol{\omega} \times (\boldsymbol{p}_i - \boldsymbol{o}) + \dot{r}(\boldsymbol{p}_i - \boldsymbol{o}) \tag{3.2}$$

By stacking the velocities of all reference points, we can express the relationship between the sphere motion and the reference points' motion in matrix form:

$$\dot{\boldsymbol{p}} = \boldsymbol{A}\boldsymbol{v}_{\text{obj}} \tag{3.3}$$

where $\boldsymbol{v}_{\text{obj}} = [\dot{\boldsymbol{o}}^T, \boldsymbol{\omega}^T, \dot{r}]^T \in \mathbb{R}^7$ represents the generalized velocity of the virtual object. The matrix $\boldsymbol{A} \in \mathbb{R}^{3N\times 7}$ is the *interaction matrix*, constructed as follows:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{I}_3 & -[\boldsymbol{p}_1 - \boldsymbol{o}]_\times & (\boldsymbol{p}_1 - \boldsymbol{o}) \\ \vdots & \vdots & \vdots \\ \boldsymbol{I}_3 & -[\boldsymbol{p}_N - \boldsymbol{o}]_\times & (\boldsymbol{p}_N - \boldsymbol{o}) \end{bmatrix} \tag{3.4}$$

Here, $\boldsymbol{I}_3$ is the 3×3 identity matrix, and $[\cdot]_\times$ denotes the skew-symmetric matrix operator for the cross product. This matrix relates the task-space velocities of the reference points to the deformation of the object.

3.3 Retargeting control law

Our retargeting strategy relies on mapping the velocity of the virtual object from the human domain to the robot domain. In the original formulation by Gioioso et al. [2], the human hand velocity is derived analytically from the synergy inputs.

Let $\boldsymbol{z} \in \mathbb{R}^{n_z}$ be the vector of synergy activation coefficients. The human joint velocities $\dot{\boldsymbol{q}}_h$ can be expressed as:

$$\dot{\boldsymbol{q}}_h = \boldsymbol{S}\dot{\boldsymbol{z}} \tag{3.5}$$

where $\boldsymbol{S} \in \mathbb{R}^{n_{q_h}\times n_z}$ is the *synergy matrix* mapping low-dimensional inputs to the full joint space. Consequently, the velocity of the human reference points $\dot{\boldsymbol{p}}_h$ would be computed as:

$$\dot{\boldsymbol{p}}_h = \boldsymbol{J}_h\dot{\boldsymbol{q}}_h = \boldsymbol{J}_h\boldsymbol{S}\dot{\boldsymbol{z}} \tag{3.6}$$

where $\boldsymbol{J}_h \in \mathbb{R}^{3N_h\times n_{q_h}}$ is the human hand Jacobian.

However, in our specific architecture, the reconstruction of the human hand pose is performed by a neural network [4], which directly outputs the full joint configuration $\boldsymbol{q}_h$ of the virtual human hand in Unity. Since the full pose is known at every frame, computing the analytical Jacobian $\boldsymbol{J}_h$ and the synergy matrix $\boldsymbol{S}$ explicitly is unnecessary. Instead, we obtain the reference point velocities $\dot{\boldsymbol{p}}_h$ via *numerical differentiation* of the tracked points in the virtual environment:

$$\dot{\boldsymbol{p}}_h(t) \approx \frac{\boldsymbol{p}_h(t) - \boldsymbol{p}_h(t-\Delta t)}{\Delta t} \tag{3.7}$$

This approach allows us to bypass the complexity of modeling the human kinematic chain analytically while ensuring that the motion fed into the retargeting algorithm accurately reflects the reconstructed hand pose.

Once $\dot{\boldsymbol{p}}_h$ is obtained, we compute the generalized velocity $\boldsymbol{v}_{\text{obj},h}$ of the human virtual sphere by inverting the interaction matrix. Since $\boldsymbol{A}_h$ is typically tall (more reference points than object DoFs), we us the *Moore-Penrose pseudoinverse* $\boldsymbol{A}_h^{\#}$:

$$\boldsymbol{v}_{\text{obj},h} = \boldsymbol{A}_h^{\#}\dot{\boldsymbol{p}}_h \tag{3.8}$$

Next, we map this motion to the robotic domain using a scaling matrix $\boldsymbol{K}_c \in \mathbb{R}^{7\times 7}$:

$$\boldsymbol{v}_{\text{obj},r} = \boldsymbol{K}_c\boldsymbol{v}_{\text{obj},h} \tag{3.9}$$

The matrix $\boldsymbol{K}_c$ scales the translation and radial growth components of the object velocity by the factor k_{sc}, while leaving the angular velocity unchanged (which is independent of size):

$$\boldsymbol{K}_c = \begin{bmatrix} k_{sc}\boldsymbol{I}_3 & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times1} \\ \boldsymbol{0}_{3\times3} & \boldsymbol{I}_3 & \boldsymbol{0}_{3\times1} \\ \boldsymbol{0}_{1\times3} & \boldsymbol{0}_{1\times3} & k_{sc} \end{bmatrix} \tag{3.10}$$

We then compute the target velocity for the robot reference points $\dot{\boldsymbol{p}}_{r,\text{des}} = \boldsymbol{A}_r \boldsymbol{v}_{\text{obj},r}$. To track these points, we use the robot's differential kinematics equation $\dot{\boldsymbol{p}}_r = \boldsymbol{J}_r \dot{\boldsymbol{q}}_r$, where $\boldsymbol{J}_r$ is the Jacobian matrix mapping robot joint velocities $\dot{\boldsymbol{q}}_r$ to end-effector velocities $\dot{\boldsymbol{p}}_r$.

The final control law for the robot joint velocities is obtained by inverting this equaton:

$$\dot{\boldsymbol{q}}_r = \boldsymbol{J}^{\#}_{r,\text{DLS}} \dot{\boldsymbol{p}}_{r,\text{des}} \tag{3.11}$$

where $\boldsymbol{J}^{\#}_r$ is the *Damped Least Squares* (DLS) inverse of the robot Jacobian, which ensures numerical stability even when the robot is near singular configurations.

3.4 Redundancy resolution

In those scenarios where dexterous robotic hands are employed, one can exploit their kinematic redundancy to optimize their internal configuration while performing grasping tasks.

To achieve this, we can augment the control law derived in Eq. 3.11 with a secondary objective using the *null-space projection* method: we exploit the null space of the Jacobian to perform secondary tasks without affecting the primary goal of tracking the virtual sphere motion. The modified control law becomes:

$$\dot{\boldsymbol{q}}_r = \underbrace{\boldsymbol{J}^{\#}_{r,\text{DLS}} \dot{\boldsymbol{p}}_{r,\text{des}}}_{\dot{\boldsymbol{q}}_{r,\text{primary}}} + (\boldsymbol{I} - \boldsymbol{J}^{\#}_{r,\text{DLS}} \boldsymbol{J}_r)\dot{\boldsymbol{q}}_0 \tag{3.12}$$

The term $(\boldsymbol{I} - \boldsymbol{J}^{\#}_{r,\text{DLS}} \boldsymbol{J}_r)$ projects an *arbitrary* velocity vector $\dot{\boldsymbol{q}}_0$ into the null space of the primary task. This allows us to define $\dot{\boldsymbol{q}}_0$ as the gradient of a performance criterion $H(\boldsymbol{q}_r)$ designed to keep the joints of the robotic hand away from the mechanical limits. Following standard kinematic control theory, we utilize the *joint range* availability function:

$$H(\boldsymbol{q}_r) = \frac{1}{2N} \sum_{i=1}^{N} \left(\frac{q_i - \bar{q}_i}{q_{i,\max} - q_{i,\min}} \right)^2 \tag{3.13}$$

where N is the number of joints, $q_{i,\max}$ and $q_{i,\min}$ are the upper and lower limits, and $\bar{q}_i$ is the midpoint of the range of joint i.

The secondary velocity task is defined as the steepest descent direction of this function:

$$\dot{\boldsymbol{q}}_0 = -\eta \nabla_{\boldsymbol{q}} H \tag{3.14}$$

where η is a positive scalar gain that regulates the influence of the secondary task, and the i-th component of the gradient is:

$$\frac{\partial H}{\partial q_i} = \frac{1}{N} \frac{q_i - \bar{q}_i}{(q_{i,\max} - q_{i,\min})^2} \tag{3.15}$$

This formulation normalizes the error, ensuring that joints with smaller ranges are prioritized (pushed harder towards the center of their range) comapred to joints with larger ranges.

For underactuated grippers with fewer degrees of freedom than the task requirements, the null space term naturally vanishes or has no effect. In these cases, the pseudoinverse in Eq. 3.11 provides the least-squares solution that minimizes the error between the desired and actual sphere motion.

Chapter 4

Experimental setup

4.1 System overview

4.1.1 Hardware

4.1.2 Software

4.2 Input layer

4.3 Software

4.4 Robotic models

Chapter 5

Results

Chapter 6

Conclusion

6.1 Summary

6.2 Future work

Bibliography

[1] M. Santello, M. Flanders, and J. F. Soechting, "Postural hand synergies for tool use," *The Journal of Neuroscience*, vol. 18, no. 23, pp. 10 105–10 115, 1998.

[2] G. Gioioso, G. Salvietti, M. Malvezzi, and D. Prattichizzo, "Mapping synergies from human to robotic hands with dissimilar kinematics: An approach in the object domain," *IEEE Transactions on Robotics*, vol. 29, no. 4, pp. 825–837, 2013.

[3] *Weart website*, https://weart.it/, Accessed: 2025-06-13.

[4] C. L. Primiceri, S. Trovalusci, D. I. Bubenek Turconi, and G. Pagano, *Motion retargeting on a human hand model*, M.Sc. Project Report, Sapienza University of Rome, Code available at: https://github.com/serenatrovalusci/Hand-Motion-Reconstruction-from-Minimal-Inputs-through-Latent-Space-Learning, Jun. 2025.