This article has been accepted for publication in IEEE Robotics and Automation Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LRA.2026.3723281

Towards Shared Control in Clutter

Miriam Welser[1], Gabriel Quere[1], Annette Hagengruber[1], Jörn Vogel[1], Freek Stulp[1], Andreas Dömel[1]

***Abstract*—Assistive robots can enable people with motor impairments to complete everyday tasks independently, offering active, semi-autonomous support. However, providing correct and seamless assistance in more cluttered multi-task environments remains a key open challenge. To address this challenge, we introduce a task-space-dependent *clutteredness* metric that quantifies scene complexity considering all tasks in the scene. This metric allows us to adapt shared control skills based on the scene's *clutteredness* with our new skill adaptation method *ShaCC* (Shared Control in Clutter). *ShaCC* comprises two steps: first, we optimize task space parameters to reduce task space overlap (*Shrinking*), and second, provide early orientation support towards the whole scene (*Merging*). We apply *ShaCC* to the Shared Control Template (SCT) framework for experimental evaluation. Our results show that the *clutteredness* metric is an expressive measure of scene complexity. Further, the task-aware skill representation of *ShaCC* improves assistance in multi-task environments with higher success rates and smoother task switching during manipulation. We show this in simulation with up to ten tasks per scene, improving the success rate by 45% on average, and validate *ShaCC* in a pilot study with three users on the assistive robot MAYA, achieving 100% task success in a real-world scene with five tasks.**

***Index Terms*—Physically Assistive Devices, Human-Centered Robotics, Intention Recognition, Shared Control, Cluttered Environments**

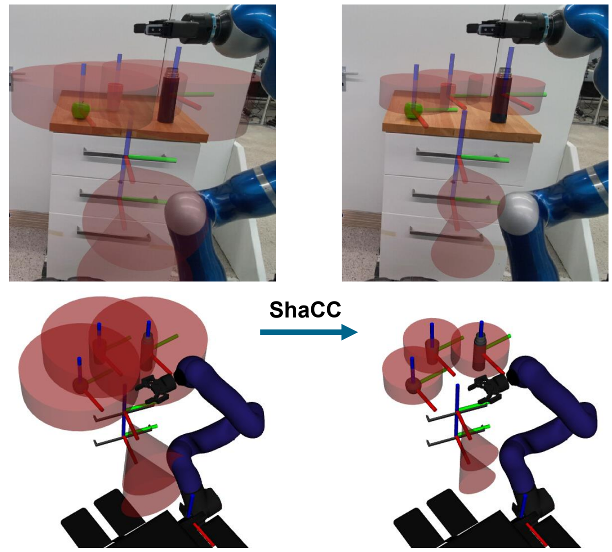


Fig. 1: Illustration of our task-space-aware *ShaCC* method on the assistive robot EDAN in a scene with five tasks. Shown is the camera view (top) and the visualization in rviz (bottom) with the task spaces highlighted in red. **Left:** SCT skills using the baseline Shared Control Template (SCT) framework. **Right:** SCT skills using *ShaCC*.

I. Introduction

Assistive robots aim to empower people with motor impairment to perform activities of daily living independently. An underlying shared control framework lets users interact with the environment effortlessly despite the system's high complexity and the use of a low-dimensional input device. Correct and seamless assistance, especially in gripper orientation, reduces mental load, increases acceptance, improves usability, and provides greater social value [1]. In contrast, incorrect assistance based on an incorrect user intent prediction degrades performance, acceptance, and trust [2], and can be worse than no assistance [3].

Providing assistance towards the correct target becomes particularly challenging in cluttered environments with multiple possible target objects and tasks. State-of-the-art methods tend to provide incorrect assistance in these environments unless querying the user directly about their target when unsure [4]. One example of a common, yet challenging multi-target scene is three targets standing in a triangle, each of which can be picked from the side (see table-top scene in Fig. 1) [5]. Reaching the mug standing behind becomes especially difficult with a geometrically explicit skill representation, which does not take other targets in the environment into account. Thus, we introduce *ShaCC* (Shared Control in Clutter), a method that considers all tasks of a scene in its skill representation.

Consequently, the following question arises: when should we adapt shared control skills to assist seamlessly in cluttered scenes? To decide when to apply skill adaptation, we identified the need to assess scene complexity, improving the comparability between scenes. We claim that the task space is an important indicator for the feasibility of a task since object distances are less expressive than how these objects could be approached when solving a task. Thus, our method uses skills with explicit task space representations, as in [6]–[9], to evaluate the complexity of a scene, computing the *clutteredness* metric based on all tasks.

This metric reflects the complexity of a scene with potentially multiple targets and skills based on explicit task space representations. Uncluttered environments with sparse targets and tasks exhibit low *clutteredness*, for which state-of-the-art shared control assistance is likely to succeed. High *clutteredness* means that task spaces are not well separated, but intersecting and ambiguous in their representation. Thus, adaptation is required only when the *clutteredness* threshold of zero is exceeded to improve assistance in cluttered scenes.

The second question addresses the skill adaptation: how can we provide seamless assistance in cluttered environments with

Manuscript received: February 20, 2026; Revised June 03, 2026; Accepted July, 18. This paper was recommended for publication by Editor Sonia Chernova upon evaluation of the Associate Editor and Reviewers' comments. This work was partially supported by the Bavarian Ministry of Economic Affairs, Regional Development and Energy, within the project "SMiLE-AI" (VLR-2506-0002).

[1]All authors are with the German Aerospace Center (DLR), Institute of Robotics and Mechatronics (RMC), Münchener Str. 20, 82234 Weßling, Germany. Corresponding author: miriam.welser@dlr.de

Digital Object Identifier (DOI): see top of this page.

This article has been accepted for publication in IEEE Robotics and Automation Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LRA.2026.3723281

multiple targets and tasks? We extend skill representations with explicit task spaces with *ShaCC* (Shared Control in Clutter) to include all tasks of a scene. First, this adapts the skills to the current scene (*Shrinking*), and second, introduces a *meta skill* to provide early orientation support (*Merging*).

We apply *ShaCC* to the state-of-the-art Shared Control Template (SCT) framework [8] for validation and experimental evaluation. SCT skills support precise positioning and provide the end-effector orientation in a task while preserving user control over the principal task dimensions. This framework has shown great usability in various experiments and demonstrations [8], [10], [11]. However, they all involve only a few targets and tasks in uncluttered environments. Thus, we evaluate *ShaCC* applied to the SCT framework on cluttered scenes in simulation and demonstrate it on the real robot. First, we assess 12 scenes with varying task space constraints and clutteredness in simulation (see Fig. 4) to provide intuition about the experimental results. Then, we sample randomized scenes with five to ten tasks to compare our method with the baseline in more cluttered scenes. On the robot, we show an exemplary scene with five tasks in a kitchen environment in a small pilot study (see Fig. 1).

This work seeks to contribute to the development of more effective shared control methods that can handle complex, cluttered real-world environments and provide users with intuitive and transparent assistance. In summary, our contributions are:

1) We introduce a task-space-aware *clutteredness* metric to measure the complexity of multi-target scenes.
2) We introduce *ShaCC* with a task-aware, geometrically explicit, and transparent skill representation for seamless interaction in multi-target environments
3) We integrate *ShaCC* into the SCT framework.
4) We evaluate *ShaCC* applied to the SCT framework in simulation and demonstrate it on the real robot.

II. Related Work

The use of explicit information gathering [4] or the selection of tasks in a GUI [12] can overcome the challenges of user intent inference in cluttered environments, such as incorrect or late assistance [4]. However, empirical studies have shown that users prefer assistance based on motion commands for seamless interaction and simplicity rather than explicit queries about their intention, which can be disturbing and irritating [1], [3], [13]. Based on the user's motion commands, assistance with the Shared Control Template (SCT) framework [8] has shown great usability not only in basic pick and place, but also in more complex and contact-rich tasks such as pouring water and opening doors [10], and even competition setups [11], all of which include only a few and sparse tasks in a scene. SCT skills are defined relative to an object. They are temporal- and spatial-invariant, enhancing both transparency and users' trust [8]. Therefore, we choose the SCT framework (see Section III) as the baseline for this first validation of *ShaCC*.

The challenge of providing robotic assistance in cluttered environments is reflected in the limited attention given to scenes that contain several targets positioned in close proximity: instead, most of the existing work on multi-target scenes focuses on related issues such as assistive control policies, active intent disambiguation, or grasping of unknown objects [1], [3], [5], [12], [14]–[16]. Consequently, systematic evaluations of real-world multi-target scenes remain largely underexplored. We aim to bridge the gap towards cluttered environments with our task-aware skill adaptation method.

Currently, the main limitation of geometrically explicit representations such as SCT skills [8] or Capture Envelopes [7] is their context-agnostic formulation that results in undesired behavior like reachability problems and wrong assistance in cluttered environments with multiple possible targets and tasks. Therefore, we propose our skill adaptation method *ShaCC* with its *Shrinking* and *Merging* steps that uses information about the *clutteredness* of a scene. It provides scene-dependent skill representations that enhance shared control assistance based on explicit task spaces in multi-target environments.

Some approaches for measuring the complexity of multi-target and cluttered scenes exist in applications such as navigation, computer vision, or autonomous robotic bin picking or grasping. These include counting objects [17], density and position of obstacles [18]–[20], object isolation and clutter metrics based on image segmentation [21], [22], or overlap and occlusions of bounding boxes in images which are used mainly in bin picking [23]–[25]. Another approach uses the 3D position of objects in a collaborative task to calculate a *clutteredness* measure based on entropy, i.e. comparing their distribution in the scene with a uniform distribution [26].

However, none of these approaches explicitly incorporates task space constraints required for object manipulation, nor are they tailored to shared control scenarios. We tackle both points by introducing a novel *clutteredness* metric based on the task space constraints of all tasks involved. For now, we omit obstacles and manipulability concerns in our method which have already been addressed in [27]. Instead, we focus on the target objects and their possibly intersecting tasks, which is a yet unexplored topic and defines the scope of this paper.

III. Background – Shared Control Templates

The *Shared Control Template* (SCT) framework provides object-centric skills that support users during task execution. Users usually control the end-effector position p_{ee}, while the end-effector orientation R_{ee} is provided by the system, as controlling it is difficult and unintuitive.

SCT skills are finite-state machines (FSMs) that define a set of states with input mappings, active constraints, state-specific parameters, and a set of transitions [8]. The *input mapping* determines how the user's motion command, typically via a low-dimensional 2D or 3D input device, is mapped to the end-effector motion, while *active constraints* restrict the end-effector pose relative to the object [8]. Together, both result in a *static, object-centric geometric task space* as is done with task space regions (TSRs) [6].

We formalize shared control assistance with SCT skills as follows: each state explicitly defines the task space $s \subseteq \mathrm{SE}(3)$ of the end-effector. The task space contains a set of allowed end-effector positions described with the *position space* $p \in \mathbb{R}^3$, for example, as a cone or a cylinder. For each

This article has been accepted for publication in IEEE Robotics and Automation Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LRA.2026.3723281

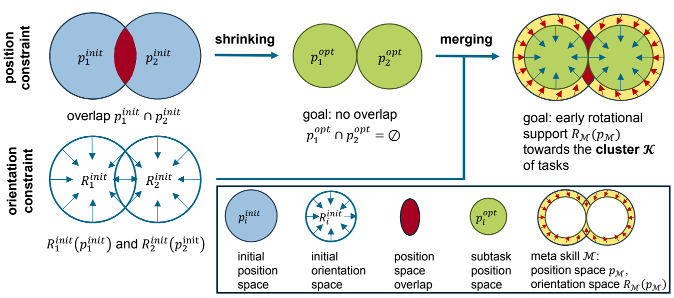


Fig. 2: *ShaCC* shown for an example scene with two tasks. Both include a disk as *position space* $p_i \in \mathbb{R}$ and a point-to-target *orientation space* $R_i(p_i) \in SO(3)$. The initial *position spaces* p_1^{init}, p_2^{init} are overlapping (marked red) leading to a *clutteredness* of $C(p_1^{\text{init}}, p_2^{\text{init}}) > 0$. *Shrinking* optimizes the *position space* parameters (radii of the disks) to reduce the *clutteredness*, while *Merging* provides early orientation support $R_{\mathcal{M}}(p_{\mathcal{M}})$ within the *meta skill position space* $p_{\mathcal{M}}$.

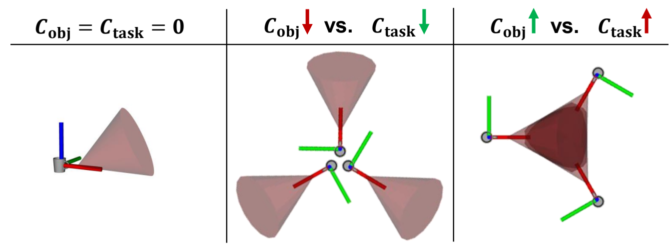


Fig. 3: Example scenes with coffee cups to be grasped by their handle using cones as position constraints. Here, *object clutteredness* and *task clutteredness* are clearly different. **Left:** Low *task* and *object clutteredness* due to the single task. **Center:** Low *task clutteredness* with 0% task space overlap while high *object clutteredness*. **Right:** High *task clutteredness* with significant task space overlap while low *object clutteredness*.

position within the *position space*, there is a *orientation space* $R(p) \in \mathrm{SO}(3)$, often reduced to a single orientation that the system provides to ease task execution with a low-dimensional input device. In combination, the end-effector pose (p_{ee}, R_{ee}) is constrained in the task space written as

$$s := \left\{ (p_{ee}, R_{ee}) \in \mathbb{R}^3 \times \mathrm{SO}(3) \;\middle|\; \begin{array}{l} p_{ee} \in p, \\ R_{ee} \in R(p_{ee}) \end{array} \right\} \subseteq \mathrm{SE}(3).$$

The SCT framework decouples *user intent inference* from *shared control assistance*. First, *user intent inference* decides whether to activate an SCT skill based on the distance of the end-effector to the target and to the *position space* of its first state using predefined thresholds [10]. When the end-effector enters the first state *position space* of an SCT skill, the skill becomes active because this region typically defines the valid entrance to successfully perform a task. For more details on *shared control assistance* with SCT skills, including end-effector projection, skill definition, and implementation, we refer to the literature [8].

IV. Methodology

We introduce multiple steps to enhance shared control based on skills with geometrically explicit task spaces for cluttered, multi-target environments. Our skill adaptation method *ShaCC* comprises three parts: (1) a *clutteredness* metric to measure scene complexity, (2) *Shrinking* to optimize task spaces, and (3) *Merging* to provide early orientation support. We propose that skills with explicit task space constraints should adapt depending on the *clutteredness* of the scene as shown in Fig. 2.

A. Clutteredness Metric

The explicit geometrical definition of task spaces provides useful information about task feasibility and scene complexity for manipulation. Therefore, *clutteredness* can be defined not only with respect to object distances as *object clutteredness*, but also with respect to task spaces as *task clutteredness*:

- *Object clutteredness* measures how close objects are placed to each other in a scene, considering their position, orientation, geometric shape, and size.
- *Task clutteredness* measures how well spaces for completing a specific task with an object are separated.

This distinction becomes particularly relevant when task spaces differ from object proximity alone, as illustrated in Fig. 3. Here, the possible tasks include grasping one of three coffee cups by the handle. The single-object scene on the left shows a single task, which implies zero *clutteredness*. In the center, the objects are placed much closer to each other, yet their task spaces are well separated, in contrast to the overlapping task spaces on the right. Therefore, we define *clutteredness* in terms of *task clutteredness*, as it more directly captures the feasibility of robot manipulation in a given scene.

Because users typically control the position of the end-effector, we consider only the *position space* for *clutteredness*. Formally, we compute the overall *clutteredness* C by adding the pairs of intersecting volumes $V(p_i \cap p_j)$ between all $n(n-1)/2$ combinations of *position spaces* p_i and p_j. We set this in relation to the sum of all *position space* volumes $V(p_i)$. More *position space* overlap is a direct measure for high *clutteredness* because it reflects how much space is occupied by more than one task, which makes the skill representation ambiguous in this region. On the other hand, *clutteredness* is zero if all *position spaces* are well separated. The metric counts areas with more than two intersecting *position spaces* more than once, so cluttered regions with multiple overlaps further increase the *clutteredness*.

$$C(p_1, \ldots, p_n) = \frac{\sum_{i=1}^{n} \sum_{j>i}^{n} V(p_i \cap p_j)}{\sum_{i=1}^{n} V(p_i)} \tag{1}$$

B. Shrinking - Task-Aware Skill Representation

For non-zero *clutteredness* $C > 0$, we adapt the skills applied on the current scene.

Formally, let $\mathcal{P}^{\text{init}} = \{p_1^{\text{init}}, \ldots, p_n^{\text{init}}\}$ denote the set of all object-centric *position spaces*, where each p_i^{init} is an initial *position space*. We aim to minimize *clutteredness*, so that $p_i^{\text{opt}} \cap p_j^{\text{opt}} \to \emptyset \quad \forall i,j$ which implies minimal *clutteredness* $C(p_1^{\text{opt}}, \ldots, p_n^{\text{opt}}) \to 0$, while strictly not exceeding parameter limits $q_k^{\text{min}} \leq q_k^{\text{opt}} \leq q_k^{\text{init}} \quad \forall k$ that are derived from the skill definition and ensure feasibility. For example, q_k could be the radius of a cylinder. To prevent task failure, we set a lower bound on q_k: the minimum radius is chosen such that task execution remains feasible, e.g., avoiding collisions with the object. Additionally, we want to keep the *position spaces* as large as possible for better usability and ensure that all tasks are well represented in the scene.

We tackle this problem by simultaneously reducing the values of the parameters q_k of intersecting task spaces iteratively until one of the following conditions is met:

- the *position space* p_i no longer has intersections, or
- the lower limit of the parameter q_k is reached.

This ensures that all task spaces are optimized concurrently until zero *clutteredness* or until further down-scaling is impossible, and at the same time maintain large *position spaces*.

C. Merging - Early Orientation Support

To address the requirement for early orientation support, we introduce *meta skills* based on the current scene that adjust the end-effector orientation R_{ee} within each *cluster* of tasks. A *cluster* $\mathcal{K}$ is a set of tasks with mutually intersecting *position spaces* (directly or through a chain of intersections), while the *cluster* as a whole does not share intersections with any *position space* of any other *cluster*. Formally, a *cluster-centric meta skill* $\mathcal{M}$ is defined for a *cluster* of tasks

$$\mathcal{K}^{\text{init}} \subseteq \mathcal{P}^{\text{init}}$$

that satisfies the following properties for its *position spaces*:

1) *Internal connectivity:* For every $p_i^{\text{init}} \in \mathcal{K}^{\text{init}}$ there exists at least one $p_j^{\text{init}} \in \mathcal{K}^{\text{init}}$ with $i \neq j$ such that
$$p_i^{\text{init}} \cap p_j^{\text{init}} \neq \varnothing.$$
2) *External disjointness:* For all $p_i^{\text{init}} \in \mathcal{K}^{\text{init}}$ and all $p_k^{\text{init}} \in \mathcal{P}^{\text{init}} \setminus \mathcal{K}^{\text{init}}$ such that
$$p_i^{\text{init}} \cap p_k^{\text{init}} = \varnothing.$$

Therefore, multiple *meta skills* may exist, one per *cluster* of tasks. Since the *orientation space* $R(p)$ is usually disjoint between tasks and defines a fixed orientation at each position, two options arise: (1) giving more freedom to the user by restricting the orientation within the range of the *orientation spaces*, or (2) providing early *cluster*-level orientation support by averaging them to a *cluster*-centric orientation $R_{\mathcal{M}}$. Given the typically low-dimensional input device and the requirement for early orientation support, we chose the latter option.

Each *meta skill* $\mathcal{M}$ is active within the union of its initial *position spaces* p_i^{init} while allowing the user to move freely in the unified *position space* $p_{\mathcal{M}}$ and choose any *subtask* within the *cluster*. An *object-centric subtask* refers to the optimized version of the initial task with a *position space* p_i^{opt}, allowing one to decide for a *subtask* in closer proximity to the target object. Orientation support of the *meta skill* is active if the end-effector is within the union of initial *position spaces* p_i^{init}, but not within any final optimized *position space* p_i^{opt}:

$$p_{\mathcal{M}} \;=\; \left(\bigcup_{i=1} p_i^{\text{init}}\right) \;\setminus\; \left(\bigcup_{i=1} p_i^{\text{opt}}\right), \;\; i = 1, \ldots, n$$

We define the desired end-effector orientation $R_{\mathcal{M}}(p_{\mathcal{M}})$ of a *cluster* $\mathcal{K}^{\text{init}}$ as the weighted average orientation of all *subtasks* of $\mathcal{K}^{\text{init}}$ using quaternion averaging as described in [28]. The weights of each quaternion depend on the Euclidean distance between the end-effector position p_{ee} and the *position space* p_i^{opt} of each *subtask*, so that nearby *subtasks* have more impact and transitions between *subtasks* become smooth:

$$w_i = \frac{1/d_i^2}{\sum_{j=1}^{n} 1/d_j^2}, \quad \text{with } d_i = d(p_{ee}, p_i^{\text{opt}}) \quad \text{s.t.} \quad \sum_{i=1}^{n} w_i = 1.$$

V. Implementation

We applied our *ShaCC* method to the SCT framework for evaluation (see Section III). For implementation details of SCT, we refer to the literature [8], [10].

Whenever the system detects a world state change, we assess the *clutteredness* and apply *ShaCC*: we first compute the intersecting volumes for the given *position spaces* using boolean mesh operations as described in [29] with trimesh [30] to evaluate the scene's complexity and group the tasks into *clusters*. A non-zero *clutteredness* indicates that ambiguous task spaces exist requiring *ShaCC*. In that case, we iteratively reduce the relevant *position space* parameters within their predefined bounds, while continuously performing collision checks using the COAL algorithm [31], until it converges to minimal *clutteredness*. The full skill adaptation of *ShaCC* completes in under 0.1s for ten tasks, which is sufficient for the updates required in shared control. Finally, we create one *meta skill* for each *cluster* of tasks to provide early orientation support and smooth task switching during operation.

We implemented *ShaCC* on two assistive robotic platforms that both employ the SCT framework: EDAN [32] and MAYA. Each consists of a commercially available wheelchair platform equipped with an 8-DoF DLR light-weight robot or a custom 7-DoF robotic arm, a Robotiq 2F85 2-finger gripper, an RGB-D camera for object detection and pose estimation, and a 3-DoF joystick as an input device.

VI. Experimental Evaluation

The main experiments to test *ShaCC* were performed in simulation to ensure repeatability and comparability of different runs. In addition, we conducted a final demonstration on the real system as described in Section VII to demonstrate the applicability and usability of our method. This section presents objectives (VI-A), our study design (VI-B), experimental results (VI-C), and discusses them (VI-D).

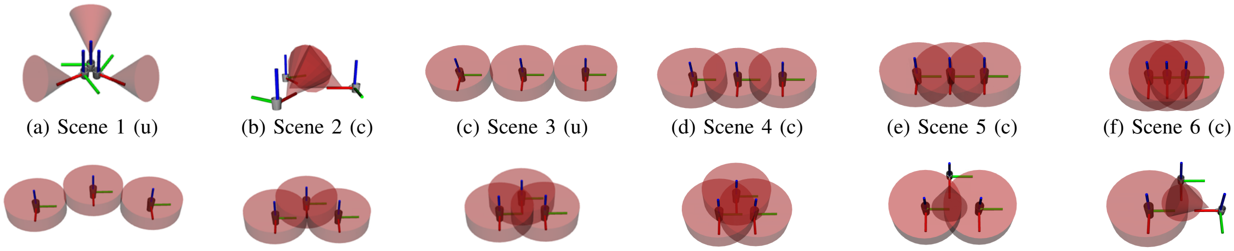
A. Objectives

The first set of experiments assesses our new *clutteredness* metric with the purpose of testing the following hypothesis.

Hypothesis 1 (H 1) *The* clutteredness *metric is negatively correlated with the success rate of reaching the intended task and assisting correctly. High* clutteredness *leads to low success rates, low* clutteredness *leads to high success rates.*

We compare our *clutteredness* metric with other distance-based measures. In this experiment, we use the baseline SCT

This article has been accepted for publication in IEEE Robotics and Automation Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LRA.2026.3723281



(a) Scene 1 (u) (b) Scene 2 (c) (c) Scene 3 (u) (d) Scene 4 (c) (e) Scene 5 (c) (f) Scene 6 (c)

(g) Scene 7 (u) (h) Scene 8 (c) (i) Scene 9 (c) (j) Scene 10 (c) (k) Scene 11 (c) (l) Scene 12 (c)

Fig. 4: Set of scenes analyzed in our first experiments. We assessed 12 manually created scenes with varying *clutteredness* (uncluttered (u) and cluttered (c)), different tasks, and asymmetry. This setup aims to provide a good intuition about the effectiveness of our method.

framework to check whether the end-effector reaches the desired positional task space successfully, and average the respective success rate per target for each scene.

The ultimate goal of this paper is to introduce and validate a method that improves shared control assistance in multi-target environments. Thus, the second set of experiments aims to test the following hypothesis.

Hypothesis 2 (H2) ShaCC *improves the baseline resulting in higher success rates and smooth task switching, i.e., less orientation change when activating a skill. Both,* Shrinking *and* Merging*, contribute to this improvement.*

To this aim, we first assess the baseline performance, not applying any part of our *ShaCC* method to SCT. Second, we apply *Shrinking* by optimizing the *position spaces* of a scene to adapt the skills. This reduces the size of the *position spaces* within (see Section IV-B). Finally, we evaluate the entire *ShaCC* pipeline including *Merging* to provide early orientation support through the *meta skill* (see Section IV-C).

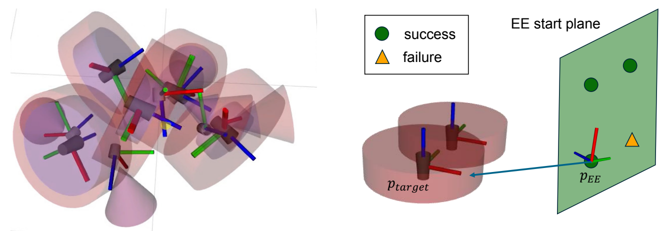
B. Study Design

We use two representative tasks for the evaluation in simulation: *pick up a coffee cup by its handle* and *pick up a mug from the side*. The corresponding SCT skills implement either a cone, which guides the end-effector towards a grasp pose, or a cylinder for axially symmetric objects, which allows the user to choose the approach direction. Both represent the most common *position spaces* used in SCT skills. *Orientation spaces* are either constraining the end-effector to point towards the target or relative to the grasp pose of the target.

For evaluation in simulation, we first designed 12 example scenes with three tasks each and varying *clutteredness* (Fig. 4) for illustration and interpretability of results, and then uniformly and Gaussian sampled randomized object poses across 20 scenes with five to ten tasks for a more abstract, but challenging setting (Fig. 5a). We further sampled 36 end-effector start positions in a vertical plane in front of the scene with arbitrary orientations (Fig. 5b). We iterated through all start poses and tasks in the scene with the goal to reach each task using the Automaton, a greedy supervised autonomy solution [33] that generates commands to move the end-effector toward the closest valid pose of a selected task.

Although our method also scales to multiple tasks per object, e.g., picking up a mug from the top or from the side, because it only cares about the actual tasks without further relation to the object itself, we decided to consider only one task per object for clarity and visualization.

We compare our *clutteredness* metric with the mean object distance d towards the center of the scene and the entropic clutteredness metric h [26] using Pearson correlation. For each experimental setup, we report the averaged scene success rates u_{scene}, and additionally, the averaged orientation delta δ_{R} when starting a skill successfully. We further calculate 95% confidence intervals via bootstrap.



(a) Randomized Scene Example (b) Experimental Setup

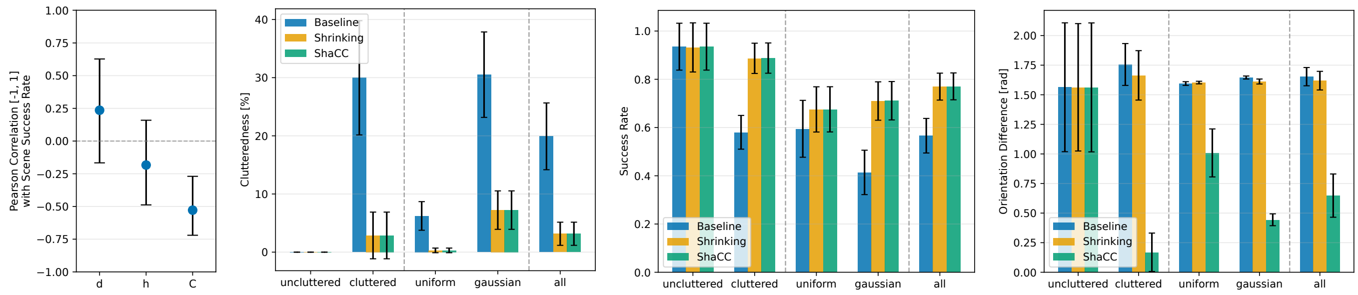
Fig. 5: (a) Example of a randomized scene for the sampled experiments showing initial (red) and optimized (blue) *position spaces*. (b) Experimental setup: the end-effector is placed at a point p_{ee} on a vertical plane in front of the scene and moves towards the desired task with p_{target} using the Automaton [33].

C. Results

The results of our experimental study in simulation are summarized in Fig. 6.

1) Clutteredness Metric: Although each of the 12 manually designed scene contains three tasks, success rates with the baseline range widely from 44% (scenes 6 and 10) to 100% (scenes 1 and 7). The sampled scenes have a lower average scene success rate of 50% ($[14\%, 96\%]$). Fig. 6a confirms that the *clutteredness* metric C exhibits a stronger correlation with the success rate (Pearson's $\rho_C = \text{-}0.53,\ 95\%\ \text{CI}_C[\text{-}0.72, \text{-}0.27]$) than with the distance-based metrics d or h whose confidence intervals include zero, indicating no correlation. Fig. 6b shows that clutteredness is on average at 30% for cluttered and Gaussian scenes, while it is zero or up to 11% for less clutter (uncluttered, uniform).

2) Shrinking: Shrinking reduces the *clutteredness* C effectively to zero for most of the manual scenes (except 10 and 11) and uniform scenes (see Fig. 6b). It reduces *clutteredness* greatly for the remaining, e.g., by -76% ($30\% \rightarrow 7\%$) for the Gaussian scenes, and by -84% ($23\% \rightarrow 4\%$) for all scenes with $C^{\text{init}} > 0\%$. This results in an average relative improvement of $+36\%$ in the success rate ($57\% \rightarrow 77\%$) as seen in



(a) Pearson Correlation (b) Clutteredness C^{final} (c) Success Rate u_{scene} [%] (d) Orientation Difference δ_{R} [rad]

Fig. 6: Evaluation of our *ShaCC* method compared to the baseline SCT and *Shrinking* on the 12 manually created scenes as shown in Fig. 4 split into the categories "cluttered" and "uncluttered", 20 randomized scenes with their sampling categories "uniform" and "gaussian", and summarized in "all". (a) Pearson correlation between success rate u_{scene} and scene complexity measures based on distance d, entropy h [26], and *clutteredness* C for the baseline experiments. (b) Final clutteredness C^{final}. (c) Mean scene success rates u_{scene}. (d) Mean difference in orientation δ_{R} at the start of a skill for successful trials. 95% confidence intervals are computed via bootstrap and shown with error bars.

Fig. 6c. Scenes that have low clutteredness with the baseline, do not show any significant improvement (uncluttered and uniform), while initially more cluttered scenes (cluttered manual and Gaussian) reveal a large improvement averaging $+61\%$ ($49\% \rightarrow 79\%$). Across all scenes with $C^{\text{init}} > 0\%$, *Shrinking* improves success rates by $+45\%$ ($52\% \rightarrow 75\%$).

3) ShaCC: Comparing the success rates of *Shrinking* with the entire *ShaCC* pipeline shows the same improvement. *Merging* only affects the orientation difference δ_R at the start of the skill, most notably in highly cluttered scenes (see Fig. 6d). The offset improves on average by -61% (1.65 rad $\rightarrow$ 0.65 rad). Across all scenes with $C^{\text{init}} > 0$, *Merging* improves δ_R by -69% (1.66 rad $\rightarrow$ 0.52 rad) compared to the baseline.

The performance of reaching the intended target is clearly visible in the success field, as shown exemplary for scene 9 in Fig. 7. Both methods succeed almost perfectly for the left- and right-mugs. The centered mug is unreachable with the baseline, as the end-effector either enters another task space first or hits the ambiguous overlap. For this example, u_{scene} improves greatly with *ShaCC* ($56\% \rightarrow 96\%$), because the reduced task spaces let the user move between the outer mugs.

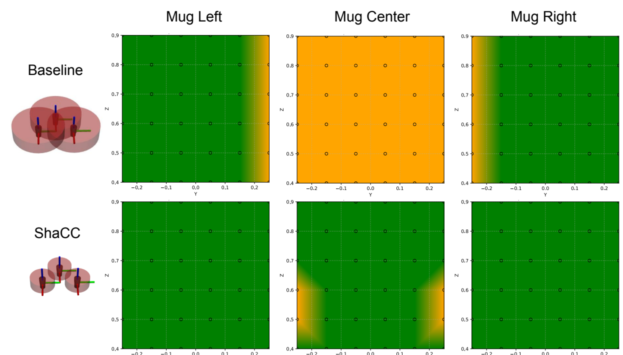


Fig. 7: Success fields for all starting points on the EE start plane (see Fig. 5b) for scene 9. It shows the success of reaching the intended targets Mug Left, Mug Center, and Mug Right with the baseline SCT (top) and our *ShaCC* method (bottom) from each point on the grid. Green areas mean success, orange failure.

D. Discussion

The results in Fig. 6 show that the *clutteredness* metric is a valid measure of scene complexity and that *ShaCC* improves the performance for cluttered scenes, while maintaining baseline performance on easy-to-reach targets.

1) Clutteredness Metric: The baseline experiments presented in Fig. 6a confirm Hypothesis 1: the *clutteredness* metric shows a strong negative correlation with the success rate of reaching the intended task (Pearson's $\rho = -0.53$, see Fig. 6a). It is more meaningful than the distance-based measures which fail most for asymmetric task spaces. For example, scene 2 with a *clutteredness* of $C = 41.8\%$, reveals a much lower success rate than scene 1 with a *clutteredness* of $C = 0\%$, even though the objects in scene 2 are farther apart. In symmetric situations, such as picking up one of three mugs from the side (e.g., scenes 3–6), distance and *clutteredness* exhibit similar correlations. Nevertheless, *clutteredness* constitutes a more general measure of scene complexity. While it captures well the ambiguity of task spaces, it does not account for reachability problems that result in low clutteredness but still low success rates, as seen in Fig. 6 for the uniform scenes.

2) ShaCC: The results in Fig. 6c and Fig. 6d confirm that *ShaCC* improves the baseline in cluttered scenes with $C^{\text{init}} > 0\%$ by improving *clutteredness* (by -84%), success rate (by $+45\%$), and orientation difference (by -69%), supporting Hypothesis 2. The success rate improvement is entirely caused by the *Shrinking* step, which is expected because *Merging* does not further change the task space of the target skill, but only the end-effector orientation within the *meta skill*. On the other hand, we see a large improvement of the orientation difference due to the *Merging* step, which will result in smoother transitions between tasks. While we cannot draw final conclusions due to the simulated environment and the greedy human model for automation and repeatability, these results validate the effectiveness of our method: *ShaCC* maintains performance in uncluttered scenes and adds value in cluttered environments in terms of success rates and orientation offset at skill start.

VII. Application on Robot

For the final robot evaluation, we extended the scene to include multi-level tasks such as picking up an apple or opening a drawer, demonstrating applicability beyond table-top environments. The scene comprised five objects (apple,

This article has been accepted for publication in IEEE Robotics and Automation Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LRA.2026.3723281

Fig. 8: Example scene with five tasks demonstrated on our assistive robot MAYA (see video) to compare *ShaCC* with the baseline SCT.

mug, thermos, two drawers), each with one associated task, as shown in Fig. 8 with an illustrative video attached.

We conducted a pilot study with three able-bodied users who are familiar with the system but new to *ShaCC*. Each session was repeated three times and began with a free exploration phase, followed by a task sequence: picking up the rear mug, placing it in front, and opening the lower drawer.

A. Results

Before performing the pick and place task, the scene contained two clusters of tasks with a final *clutteredness* that decreased from 27% with the baseline to 0.4% with *ShaCC*.

Exploring the scene was hindered by the large task spaces of the baseline, which led to early, but wrong assistance, inferring the thermos or the apple as the intended target. On the other hand, *ShaCC* activated early orientation support with the *meta skill*. This allowed users to switch between tasks in close proximity to the objects, while the orientation of the end-effector changes smoothly as seen in the attached video.

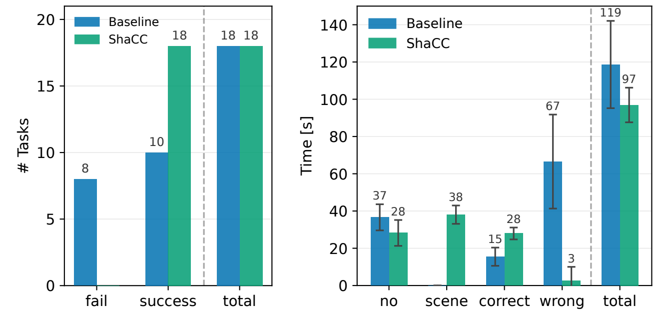
With the baseline method, eight (of nine) trials of the pick and place task failed, while the open drawer task always worked, leading to an overall success rate of 58%. With *ShaCC*, users were able to complete the entire sequence of tasks. Still, *ShaCC* saved 30% of execution time compared to the baseline SCT. More importantly, it reduced the fraction of time that the robot inferred users' intent wrongly and provided wrong assistance from 56% to 3% as shown in Fig. 9b.

After picking and placing the mug, the skills for the scene were updated, and *ShaCC* adapted them to the new scene.

B. Discussion

The impact of *ShaCC* was shown not only in simulation, but also tested on the real system in a pilot study, indicating that our method is of practical use and has the potential to improve intent prediction accuracy and execution times (see Fig. 9b). We showed that *ShaCC* can handle multiple targets in close proximity in multi-layered environments, while the baseline method fails in high clutter.

The pilot study indicates that *ShaCC* allows users to seamlessly switch tasks and select the correct one more reliably. Picking the mug required moving the end-effector between the two other objects on the surface. This consistently triggered wrong assistance with SCT, while *ShaCC* opened a corridor



(a) Task Success (b) Execution Times [s] $\pm$ std

Fig. 9: Results of the pilot study on the real robot. (a) Accumulated task success per method. (b) Averaged execution times split by the robot assistance compared to the user intent: no (default 3D translational control), scene (*meta skill* for early orientation support), and correct / wrong assistance (towards the correct / wrong goal).

to pass through. For the drawer task, all users successfully navigated around the upper drawer's task space to reach the lower one. *ShaCC* would have permitted a shorter path in front of the drawer as seen in the video, but most users appeared to default to the baseline strategy they are already familiar with.

The first validation of *ShaCC* in a pilot study indicates that users can easily reach the desired task and can even switch seamlessly between different tasks with our method. Compared to that, the baseline is more likely to activate a wrong skill because of the inadequate skill representation in a cluttered scene and even hinders users from reaching their intended target. Objective measures suggest that the robot provides better and more correct assistance in cluttered environments with our *ShaCC* method, which is in line with the requirement of correct, early, and orientation assistance [1].

VIII. Limitations and Future Work

This paper presents a first validation of *ShaCC* via simulation and a pilot study on the real robot. Below, we discuss key limitations to contextualize results and guide future research.

So far, we evaluated *ShaCC* against a single baseline using objective measures for repeatability. As future work, we plan an extensive target user study to draw more robust conclusions on usability, generalizability, and real-world impact.

Some scenes cannot reach zero clutteredness due to the parameter bounds that ensure feasibility, which leads to lower success rates especially in the highly cluttered randomized experiments. Some of the failures can be explained by the greedy user input model that ignores obstacles compared to a real user who might simply move around them or modify the scene. Introducing scene-level skills or combining *ShaCC* with non-prehensile manipulation to actively modify the environment could further improve assistance in these challenging cases.

The merging step has two limitations: the averaged orientation may not point directly towards any target, which can feel unintuitive and will be assessed in the planned user study, and opposing orientation constraints can cause a sudden flip when moving between two tasks. Since these locations are precomputable, they could serve as an indicator that neither task is intended, an interesting direction for future work.

This article has been accepted for publication in IEEE Robotics and Automation Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LRA.2026.3723281

ShaCC was developed for shared control systems with explicit task spaces, such as SCT [8], Capture Envelopes [7], or skills derived from Kernelized Movement Primitives [9], and is complementary to methods based on implicit policies such as Hindsight Optimization [1] or POMDPs [3], [4] that directly evaluate the user input. *ShaCC* is agnostic to the user input (direction) because it evaluates the end-effector position in relation to the task spaces statically. Combining those dynamic methods with *ShaCC* remains an interesting open topic.

IX. Conclusion

This paper introduced *ShaCC*, a new skill adaptation method for shared control in cluttered environments with multiple targets. First, we presented *clutteredness*, a general measure for scene complexity that can be employed with any skill representations using explicit task spaces. Second, we introduced *Shrinking* and *Merging* to adapt skills to the current scene based on the available tasks. Finally, we implemented and evaluated *ShaCC* within the SCT framework, discussed the results and outlined directions for future work.

References

[1] S. Javdani, H. Admoni, S. Pellegrinelli, S. S. Srinivasa, and J. A. Bagnell, "Shared autonomy via hindsight optimization for teleoperation and teaming," *The International Journal of Robotics Research*, vol. 37, no. 7, pp. 717–742, Jun. 2018.

[2] A. D. Dragan and S. S. Srinivasa, "Formalizing assistive teleoperation," in *Robotics: Science and Systems*, 2012. [Online]. Available: https://api.semanticscholar.org/CorpusID:77662

[3] S. Jain and B. Argall, "Probabilistic human intent recognition for shared autonomy in assistive robotics," *ACM Transactions on Human-Robot Interaction*, vol. 9, no. 1, pp. 1–23, Dec. 2019.

[4] J.-A. Yow, N. P. Garg, and W. T. Ang, "Shared autonomy of a robotic manipulator for grasping under human intent uncertainty using pomdps," in *IEEE Transactions on Robotics, vol. 40, pp. 332-350*. IEEE, 2024.

[5] A. Belsare, Z. Karimi, C. Mattson, and D. S. Brown, "Toward zero-shot user intent recognition in shared autonomy," in *IEEE/ACM International Conference on Human-Robot Interaction (HRI), 2025*, 2025.

[6] D. Berenson, S. Srinivasa, and J. Kuffner, "Task space regions: A framework for pose-constrained manipulation planning," *The International Journal of Robotics Research*, vol. 30, no. 12, pp. 1435–1460, Mar. 2011.

[7] K. Muelling, A. Venkatraman, J.-S. Valois, J. E. Downey, J. Weiss, S. Javdani, M. Hebert, A. B. Schwartz, J. L. Collinger, and J. A. Bagnell, "Autonomy infused teleoperation with application to bci manipulation," in *Autonomous Robots*, 2015.

[8] G. Quere, A. Hagengruber, M. Iskandar, and S. Bustamante, "Shared control templates for assistive robotics," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020.

[9] G. Quere, F. Stulp, D. Filliat, and J. Silvério, "A probabilistic approach for learning and adapting shared control skills with the human in the loop," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 15 728–15 734.

[10] A. Hagengruber, G. Quere, M. Iskandar, S. Bustamante, J. Feng, D. Leidner, A. Albu-Schäffer, F. Stulp, and J. Vogel, "An assistive robot that enables people with amyotrophia to perform sequences of everyday activities," in *Scientific Reports*, 2025.

[11] J. Vogel, M. Atzenhofer, G. Quere, M. Iskandar, M. Welser, F. Schiel, S. Jung, S. Bustamante, W. Friedl, W. Boerdijk, F. Stulp, A. Albu-Schäffer, and A. Hagengruber, "Shared control in assistive robotics: A cybathlon-winning approach," *Science Robotics*, vol. 10, no. 107, p. eaeb6725, 2025.

[12] E. Miller, M. Durner, M. Humt, G. Quere, W. Boerdijk, A. M. Sundaram, F. Stulp, and J. Vogel, "Unknown object grasping for assistive robotics," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024.

[13] A. D. Dragan and S. S. Srinivasa, "A policy-blending formalism for shared control," *The International Journal of Robotics Research*, vol. 32, no. 7, pp. 790–805, Jun. 2013.

[14] D. E. Gopinath and B. D. Argall, "Active intent disambiguation for shared control robots," in *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. IEEE, 2020.

[15] D.-J. Kim, R. Hazlett-Knudsen, H. Culver-Godfrey, G. Rucks, T. Cunningham, D. Portee, J. Bricout, Z. Wang, and A. Behal, "How autonomy impacts performance and satisfaction: Results from a study with spinal cord injured subjects using an assistive robot," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, no. effective, 2012.

[16] A. Gottardi, S. Tortora, E. Tosello, and E. Menegatti, "Shared control in robot teleoperation with improved potential fields," in *IEEE Transactions on Human-Machine Systems*. IEEE, 2022.

[17] M. R. Dogar and S. S. Srinivasa, "A planning framework for non-prehensile manipulation under clutter and uncertainty," *Autonomous Robots*, vol. 33, no. 3, pp. 217–236, Jun. 2012.

[18] G. T. Anderson and G. Yang, "A proposed measure of environmental complexity for robotic applications," in *2007 IEEE International Conference on Systems, Man and Cybernetics*. IEEE, 2007.

[19] Y. S. Shao, Y. Wu, L. Jarin-Lipschitz, P. Chaudhari, and V. Kumar, "Design and evaluation of motion planners for quadrotors in environments with varying complexities," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024.

[20] A. Ram, R. Arkin, K. Moorman, and R. Clark, "Case-based reactive navigation: A method for on-line selection and adaptation of reactive robotic co - systems, man and cybernetics, part b, ieee transactions on," in *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) (Volume: 27, Issue: 3, June 1997)*, vol. 27, no. 3. IEEE, jun 1997, pp. 376–394.

[21] M. Zhao, G. Zuo, S. Yu, D. Gong, Z. Wang, and O. Sie, "Position-aware pushing and grasping synergy with deep reinforcement learning in clutter," *CAAI Transactions on Intelligence Technology*, vol. 9, no. 3, pp. 738–755, Aug. 2023.

[22] N. Lu, T. Lu, Y. Cai, S. Wang, R. C. on Intelligent Robotic Systems, I. of Automation, C. A. of Sciences, Beijing, and China, "Active pushing for better grasping in dense clutter with deep reinforcement learning," in *2020 Chinese Automation Congress (CAC)*. IEEE, 2020.

[23] B. I. van Blokland and T. Theoharis, "Radial intersection count image: A clutter resistant 3d shape descriptor," *Computers & Graphics*, vol. 91, no. Section, pp. 118–128, Oct. 2020.

[24] A. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3d scenes," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*. IEEE, 1999.

[25] S. D'Avella, M. Bianchi, A. M. Sundaram, C. A. Avizzano, M. A. Roa, and P. Tripicchio, "The cluttered environment picking benchmark (cepb) for advanced warehouse automation: Evaluating the perception, planning, control, and grasping of manipulation systems," in *IEEE Robotics and Automation Magazine*. IEEE, 2024.

[26] M. Schmidt-Wolf, T. Becker, D. Oliva, M. Nicolescu, and D. Feil-Seifer, "Through the clutter: Exploring the impact of complex environments on the legibility of robot motion," in *2025 IEEE International Conference on Robotics and Automation (ICRA)*, 2024.

[27] S. Bustamante, I. Rodríguez, G. Quere, P. Lehner, M. Iskandar, D. Leidner, A. Dömel, A. Albu-Schäffer, J. Vogel, and F. Stulp, "Feasibility checking and constraint refinement for shared control in assistive robotics," *IEEE Robotics and Automation Letters*, vol. 9, no. 9, pp. 8019–8026, 2024.

[28] F. L. Markley, Y. Cheng, J. L. Crassidis, and Y. Oshman, "Averaging quaternions," *Journal of Guidance, Control, and Dynamics*, vol. 30, no. 4, pp. 1193–1197, 2007. [Online]. Available: https://doi.org/10.2514/1.28949

[29] J. M. Smith, "Towards robust inexact geometric computation," University of Cambridge, Computer Laboratory, Tech. Rep. UCAM-CL-TR-766, Dec. 2009. [Online]. Available: https://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-766.pdf

[30] D.-H. et al., "trimesh." [Online]. Available: https://trimesh.org/

[31] J. Pan, S. Chitta, D. Manocha, F. Lamiraux, J. Mirabel, J. Carpentier, L. Montaut *et al.*, "Coal: an extension of the flexible collision library," https://github.com/coal-library/coal, 2015–2024.

[32] J. Vogel, A. Hagengruber, M. Iskandar, G. Quere, U. Leipscher, S. Bustamante, A. Dietrich, H. Höppner, D. Leidner, and A. Albu-Schäffer, "Edan: An emg-controlled daily assistant to help people with physical disabilities," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 2020*. IEEE, 2020.

[33] S. Bustamante, G. Quere, K. Hagmann, X. Wu, P. Schmaus, J. Vogel, F. Stulp, and D. Leidner, "Toward seamless transitions between shared control and supervised autonomy in robotic assistance," in *IEEE Robotics and Automation Letters*. IEEE, 2021.