

Grasp Pose Generation from RGB-D Geometry and User Approach Direction

1 Project Introduction

This project develops a shared autonomy grasping layer on top of an existing teleoperation system that uses direction vectors referenced on the surface points. The hardware setup includes a 6-DoF robot arm, a 2-finger parallel gripper, and an RGB-D camera. The main goal is to reduce the user effort during the final approach and grasp alignment phase by:

- Grasp pose estimation: generate and maintain a feasible grasp pose (6-DoF) from RGB-D geometry, biased by the user’s current approach direction.
- Shared autonomy assistance: assist the final approach by stabilizing position/orientation toward that grasp pose with confidence-gated, overrideable guidance.

The project avoids training a new grasp model. Instead, it explores geometry-based synthesis or pretrained AI models.

2 System Overview

When the end-effector is close to an object, the grasping module runs the following pipeline:

1. Perception: segment the target object (e.g., plane removal + clustering).
2. Candidate generation: propose multiple grasp poses consistent with a parallel-jaw gripper.
3. Feasibility filtering: remove candidates that violate gripper constraints, inverse kinematics (IK), or collision requirements.
4. Selection: choose and maintain a single best candidate using a confidence score and temporal smoothing.
5. Shared autonomy assistance: as the user continues teleoperating, bias the end-effector pose toward the selected grasp (position + orientation) with assistance strength controlled by confidence and user override cues.

3 Grasp Pose Exploratory Directions

A core part of the project is to implement and compare different ways of generating grasp pose candidates, without training a new grasp network.

3.1 Option A: Approach-aware geometry-based grasp synthesis

This aims to generate grasp candidates directly from the object’s RGB-D geometry and the user’s current approach direction. The method:

- estimates a simple object reference frame from the point cloud (e.g., using PCA) to get a few stable directions for grasping,

- samples a small number of parallel-jaw grasp poses by placing the gripper around the object along these directions,
- biases the sampled poses toward approaches that match the user’s current approach direction (to avoid grasps that require coming from an awkward or blocked side),
- outputs a small set of diverse candidates (e.g., 5–20) before feasibility filtering.

3.2 Option B: Pretrained grasp pose generators

Instead of training, we can also use a pretrained grasp detection model to produce a large set of 6-DoF grasp candidates from depth/point clouds, then apply our own filtering and shared autonomy logic on top.

Examples to explore include:

- Contact-GraspNet (depth/point cloud to parallel-jaw grasps): <https://arxiv.org/abs/2103.14127>
- GraspNet baselines (pretrained models on large grasp datasets): <https://github.com/graspsnet/graspsnet-baseline>
- GPD-style grasp candidate generation: <https://arxiv.org/abs/1706.09911>

The contribution here is not improving the grasp network, but integrating candidate grasps into our direction vector based teleoperation behavior.

3.3 Option C: VLM/LLM-guided grasp generation via structured constraints

We can explore how to use VLM/LLMs to support grasp generation by producing structured grasp constraints that are converted into concrete poses using geometry. For example, given an image crop and a short instruction, the model can output:

- grasp region: handle vs body vs rim,
- approach preference: from the side vs from above,
- avoid regions: openings, fragile parts, clutter side.

The system then generates grasp candidates on the point cloud consistent with these constraints and runs the same feasibility checks. This makes the AI component useful and grounded, without requiring training a grasp predictor.

4 Automatic Selection and Confidence-Gated Assistance

The system maintains a shortlist of feasible candidates and automatically commits to one when it is consistently best. Key ideas:

- Confidence scoring: combine geometric grasp quality (e.g., opposing contacts, width margin), reachability/IK margin, and approach clearance into a confidence score.
- Temporal stability: apply smoothing/hysteresis so the chosen grasp does not flicker frame-to-frame.
- Assistance policy: increase assistance when confidence is high and the user’s motion is consistent; reduce assistance quickly when confidence drops or the user overrides (moves away/rotates away).

The end result is a predictable shared autonomy behavior: the robot helps more when it is sure, and stays out of the way when uncertain.

5 Evaluation Plan

We will evaluate the grasping layer using teleoperated grasping and basic pick-and-place scenarios:

- single object grasping from multiple approach directions,
- cluttered scenes with distractors,
- objects with multiple plausible grasp styles (e.g., mug handle vs body),
- placement into a constrained region after grasping (simple pick-and-place).

6 Uniqueness and Expected Contribution

This project is unique in its shared autonomy integration, not in inventing a new grasp predictor. Many grasp detection methods focus on generating grasps for autonomous execution. In contrast, this project focuses on how grasp candidates are made usable in teleoperation:

- Approach-aware grasp generation: candidates are conditioned on the operator’s approach direction, producing grasps that match how the user is already moving the robot.
- Confidence-gated assistance: assistance strength depends on feasibility and reliability signals, improving predictability.
- Training-free design with AI compatibility: the system works with geometry-only grasp synthesis, can incorporate pretrained grasp models, and can optionally use VLM/LLM outputs as structured constraints without requiring new training.

The final outcome is a reusable grasping shared autonomy module that can serve as a starting point for future vision-based shared autonomy research.

7 Recommended Papers to Read

1. Autonomy Infused Teleoperation with Application to BCI Manipulation (RSS 2015)
- <https://www.roboticsproceedings.org/rss11/p39.pdf>
2. Intent-based Task-Oriented Shared Control for Intuitive Telemanipulation
- <https://link.springer.com/content/pdf/10.1007/s10846-024-02185-1.pdf>
3. An Intent-based Task-aware Shared Control Framework for Intuitive Hands Free Telemanipulation
- <https://arxiv.org/abs/2003.03677>
4. Shared Autonomy for Intuitive Teleoperation
- <https://www.honda-ri.de/pubs/pdf/5098.pdf>
5. Assistive Control of Robot Arms via Adaptive Shared Autonomy (AIM 2024)
- <https://ras.papercept.net/images/temp/AIM/files/0378.pdf>
6. Sampling-Based Grasp and Collision Prediction for Assisted Teleoperation
- <https://arxiv.org/abs/2504.18186>