

# Supplementary Materials

Yanming Shao, Chenxi Xiao\*

October 25, 2024

## Verifying Bimanual Grasps by Geometric Primitives

Besides the experiments in the manuscripts, we additionally evaluate our proposed BimanGrasp algorithm through a series of experiments on geometric primitives. Those geometric primitives are normalized with diameter  $d = 0.32$  m (defined as the the Euclidean distance between two farthest points in the object meshes), with shapes including rods, spheres, ellipsoids, disks, and torus, as shown in Fig. 1. These geometric primitives are much larger than the graspable volume of a single *Shadow Hand* manipulator, whose knuckle width (i.e. the width from the index finger to the little finger) is 0.084 m. Therefore, these objects pose challenges for unimanual manipulation. Besides, the unique shapes of these objects present specific challenges for our grasp synthesis algorithm, as they require diverse forms of grasps [1, 2]. For instance, a rod requires a firm grip on its cylindrical handle, while ellipsoids and polygons need to properly balance their center of masses to prevent slippage. Similarly, a torus, with its donut-like shape, is most suitable to be gripped from two opposite sides. From results in Fig. 1, our BimanGrasp algorithm can obtain stable grasps for all these objects, demonstrating its versatility and effectiveness in handling diverse geometric shapes.

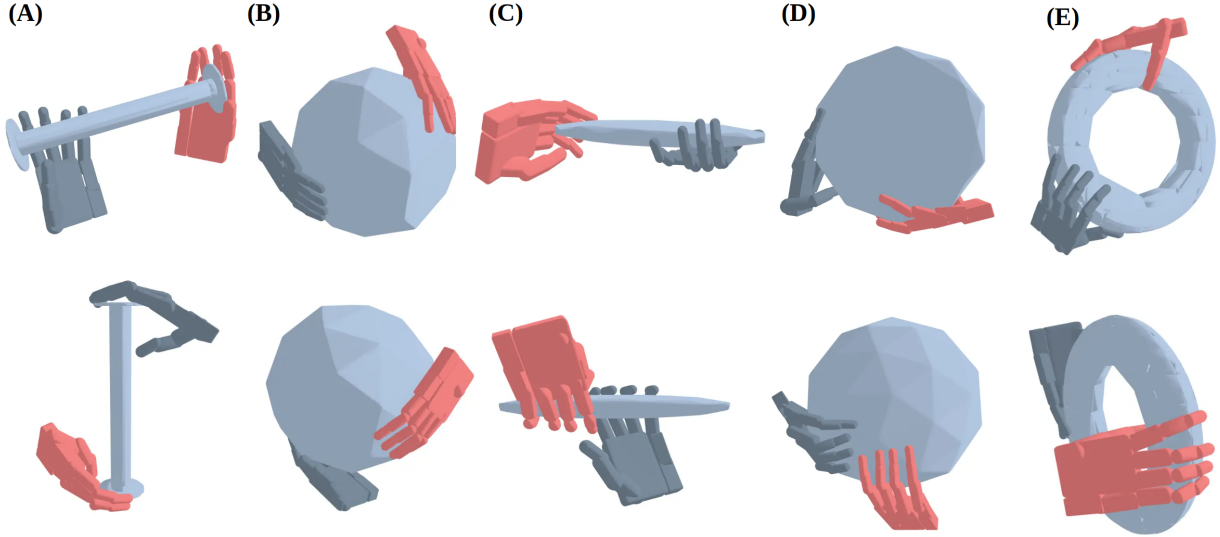


Figure 1: Visualization of the grasp poses synthesized for geometric primitives using BimanGrasp algorithm. Objects include: (A) rod, (B) polygon, (C) ellipsoid, (D) disk, (E) torus.

## Details of BimanGrasp-DDPM Implementation

We provide the parameters of our BimanGrasp-DDPM in the following table:

Table 1: Parameters of BimanGrasp-DDPM

	Parameter	Value
Dimensions	in_channels	56
	out_channels	56
	time_dim	256
	condition_dim	1024
DDPM Parameters	beta_start	0.0001
	beta_end	0.02
	num_timesteps	1000
	diffuser.steps	100
Training	loss_type	$l_1$
	num_epochs	3000
	learning_rate	0.0002

## Details of Our VLM Scorer

In Sec. IV. A paragraph 2, we used *GPT-4 Vision* as a bimanual grasp scorer. Here is the prompt we adopted (Listing. 1) (extended from [3]):

```
1 "You are a scorer for bimanual grasp poses. Please rate the quality of the
   bimanual grasps shown in the picture. Before assigning a score, consider
   the following three questions: Which part of the object does it grasp?
   WHich fingers are being used to grasp the object? Does the hand pose look
   natural and in line with typical human habits? Rate 1,2,3 (3 is highest)
   of the grasp."
```

Listing 1: Bimanual Grasp Scorer’s Prompt

2D screenshots of the bimanual grasps were used for scoring. We provide some samples of scorer’s answers, as detailed below (Listing. 2,3):

Listing 2: Answer Instance 1 (score: 2) from our VLM scorer.

```
1 Answer 1 (score: 2): The red hand is using all fingers, including the thumb, for a
   full and secure grip. The blue hand, however, appears to be using only the thumb
   and fingers without wrapping completely around the object, suggesting a less
   secured grip. The pose of the red hand appears natural and typical for handling
   such objects, providing a firm grip. However, the blue hand does not seem to grasp
   in a natural or comfortable manner which might not be efficient or safe in a real-
   world application. **Rating:** Based on the criteria, I would rate the overall
   grasp quality at a **2 out of 3**. The grip by the red hand is excellent, but the
   grasp by the blue hand is somewhat lacking in security and natural positioning
   which might affect the efficacy of the bimanual manipulation of the object."
```

Listing 3: Answer Instance 2 (score: 3) from our VLM scorer.

1 Answer 2 (score: 3): In the image, we see two hands engaging in a bimanual grasp around a bottle. The hands are grasping the body and neck of the bottle, which is typical for carrying or handling a bottle, especially if stability or pouring control is needed. The red hand is using its thumb, index, and middle finger to grasp the neck of the bottle, while the white hand is using the thumb, index, and middle finger as well, along with some engagement from the ring and little fingers to support the body of the bottle. The hand pose looks quite natural and reflect common human behavior when handling a bottle. The grasp appears secure and controlled. Given these observations, I would rate the quality of the bimanual grasp as **\*\*3\*\***. It effectively uses the fingers to stabilize and handle the object, the grasp positions are appropriate for the object type, and the pose looks natural and ergonomic.

## References

- [1] T. Feix, J. Romero, H.-B. Schmiedmayer, A. M. Dollar, and D. Kragic, "The grasp taxonomy of human grasp types," *IEEE Transactions on human-machine systems*, vol. 46, no. 1, pp. 66–77, 2015.
- [2] F. Krebs and T. Asfour, "A bimanual manipulation taxonomy," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 11 031–11 038, 2022.
- [3] Y. Liu, Y. Yang, Y. Wang, X. Wu, J. Wang, Y. Yao, S. Schwertfeger, S. Yang, W. Wang, J. Yu *et al.*, "Realdex: Towards human-like grasping for robotic dexterous hand," *arXiv preprint arXiv:2402.13853*, 2024.