Stability-driven Contact Reconstruction From Monocular Color Images

– Supplementary Material –

Zimeng Zhao Binghui Zuo Wei Xie Yangang Wang*

Southeast University, China

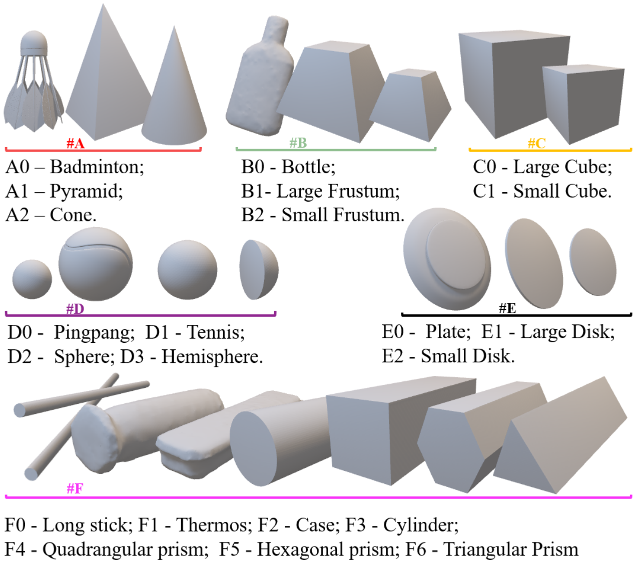

#A
A0 – Badminton;
A1 – Pyramid;
A2 – Cone.
#B
B0 - Bottle;
B1- Large Frustum;
B2 - Small Frustum.
#C
C0 - Large Cube;
C1 - Small Cube.
#D
D0 - Pingpang; D1 - Tennis;
D2 - Sphere; D3 - Hemisphere.
#E
E0 - Plate; E1 - Large Disk;
E2 - Small Disk.
#F
F0 - Long stick; F1 - Thermos; F2 - Case; F3 - Cylinder;
F4 - Quadrangular prism; F5 - Hexagonal prism; F6 - Triangular Prism

Figure 1. **Objects and corresponding IDs in each category**. There are 20 objects classified into 6 categories according to their shape, and each category contains multiple regular geometries and a multi-view reconstruction model.

A. Overview

In this supplementary document, we first introduce our interaction dataset, named CBF (Contact with Balancing Force), which contains physical attributes of hand-object and stability degree measured by the magnitude of the extra balancing force (Sec. B). More ablation studies of our sampling-based optimization are discussed in Sec. C. They were not included in the paper due to the page limit.

B. Capturing System

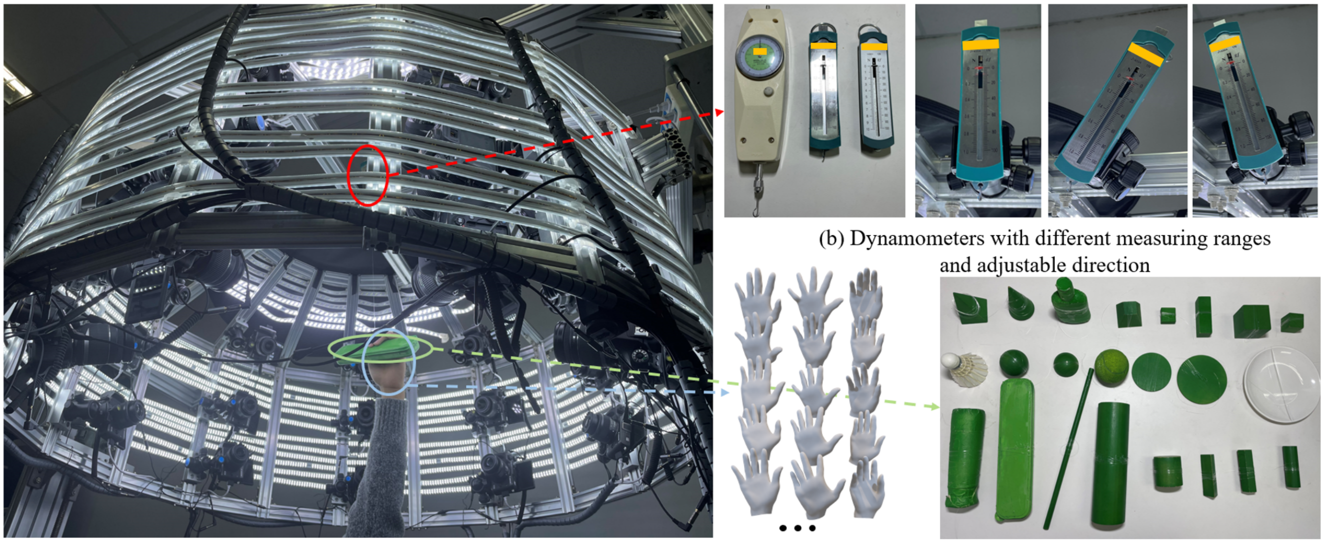
The experimental image data was captured by 25 calibrated cameras capable of capturing images. The capturing signal of the camera system is triggered when the indication of the dynamometer and the hand object interaction state is no longer changed. During the capture process, the multi-view system records the 3D state of the hand object, and the hand-object contact includes the stable states and the unstable states. As shown in Fig. 2, the additional external force needed to maintain the object balance is recorded by a dynamometer with a suitable range suspended from the top of the system.

Objects. As shown in Fig. 1, there are 20 objects classified into 6 categories according to their shape: cones, prisms, cubes, spheres, disks, and columns. On average, each category contains multiple regular geometries and one everyday object. Among them, the regular geometries are all made of the same wood. Most objects are painted green to facilitate segmentation from the image. The object meshes were acquired in three ways:

- Scan reconstruction. The meshes of B0, F1, F2 were created using the multi-view reconstruction method. In this step, an additional texture [4] is applied to the surface of the object;
- CAD modeling. The meshes of the regular geometries were created in the modeling software based on the measured sizes;
- Online Searching. The meshes of A0, D0, D1, and E0 were first downloaded from Thingiverse [2]. After that, the size of each mesh was fine-tuned according to the size of our real object.

The physical attributes of all interacted objects are recorded in Tab. 1. The size of the object corresponds to the length, width, and height of the object mesh bounding box in the rest pose. The mass of each object was weighed by a high-precision digital electronic scale. The friction between each object and flat subject skin was measured by the dynamometer in a dry environment. Each physical quantity

*Corresponding author. E-mail: yangangwang@seu.edu.cn. This work was supported in part by the National Natural Science Foundation of China (No. 62076061), the “Young Elite Scientists Sponsorship Program by CAST” (No. YES20200025), and the “Zhishan Young Scholar” Program of Southeast University (No. 2242021R41083).



(b) Dynamometers with different measuring ranges and adjustable direction

(a) Capturing System

(c) Personalized hands

(c) Objects attached with lines

Figure 2. **Capturing System**. Multi-view system with balancing force measuring devices.

Objects (ID)	Size (mm)	Mass (g)	Friction
A0	$(67, 67, 72)$	5.1	0.010
A1	$(49, 49, 77)$	46.7	0.659
A2	$(50, 50, 77)$	0.636	0.673
B0	$(64, 28, 131)$	123.9	0.641
B1	$(48, 48, 77)$	37.5	0.647
B2	$(30, 30, 48)$	18.3	0.621
C0	$(50, 50, 50)$	87.1	0.689
C1	$(29, 29, 29)$	18.9	0.423
D0	$(40, 40, 40)$	3.0	0.086
D1	$(66, 66, 66)$	56.2	0.854
D2	$(48, 48, 48)$	44.8	0.670
D3	$(48, 48, 24)$	23.4	0.513
E0	$(152, 152, 20)$	95.8	0.412
E1	$(100, 100, 3)$	13.7	0.342
E2	$(80, 80, 2)$	6.4	0.336
F0	$(10, 10, 300)$	10.9	0.381
F1	$(50, 50, 162)$	39.8	0.653
F2	$(30, 58, 245)$	67.8	0.443
F3	$(23, 23, 75)$	414.2	0.687
F4	$(25, 25, 76)$	37.1	0.691
F5	$(29, 29, 75)$	27.2	0.748
F6	$(23, 23, 75)$	15.9	0.722

Table 1. **Object physical attributes.** Our dataset provides physical attributes for each object, including object size, mass, and friction coefficient between the object and the hand (average from multiple measurements).

was measured multiple times and averaged to minimize error. When preparing training data, the radii of the object ellipsoid correspond to the size of the object bounding box in its rest pose. The 6 DoF pose of the ellipsoid is consistent with the object transformation from the rest pose to the pose in images, which is annotated in most existing datasets. Although similar radii or symmetry in an object mesh may lead to confusion into those annotations, this risk seems to be common to the pipeline represented object pose with an oriented bounding box.

Subjects. A total of 20 people were invited to participate in the production of the dataset, including 10 males and 10 females. All the involved subjects agreed to the release of the dataset, and their consent forms are on the last page of this supplementary. The friction coefficient was remeasured for each subject, and column 4 of Tab. 1 reflects the average value. As shown in Fig. 2(c), the personalized hand parameters of the subject are pre-optimized by the multi-view system. In the subsequent hand-object state reconstruction, only the pose parameters of the personalized hand are optimized.

Markless Capture. Each hand-object interaction state is observed through 25 calibrated cameras. Most of these camera models are Canon EOS M6 with high resolution. An additional camera facing the dynamometer was used to record the magnitude of the equilibrium force. All these cameras were connected by the common audio cable for synchronization. The trigger signal was sent out of the system by a high-speed wireless remote controller. For each RGB frame, we performed instance segmentation to get hand mask and object mask manually. Then we ran the SRHandNet [5] to locate the key-points $\boldsymbol{x}^{\star} \in \mathbb{R}^{2\times 21}$. Using the relative position relationship between the object

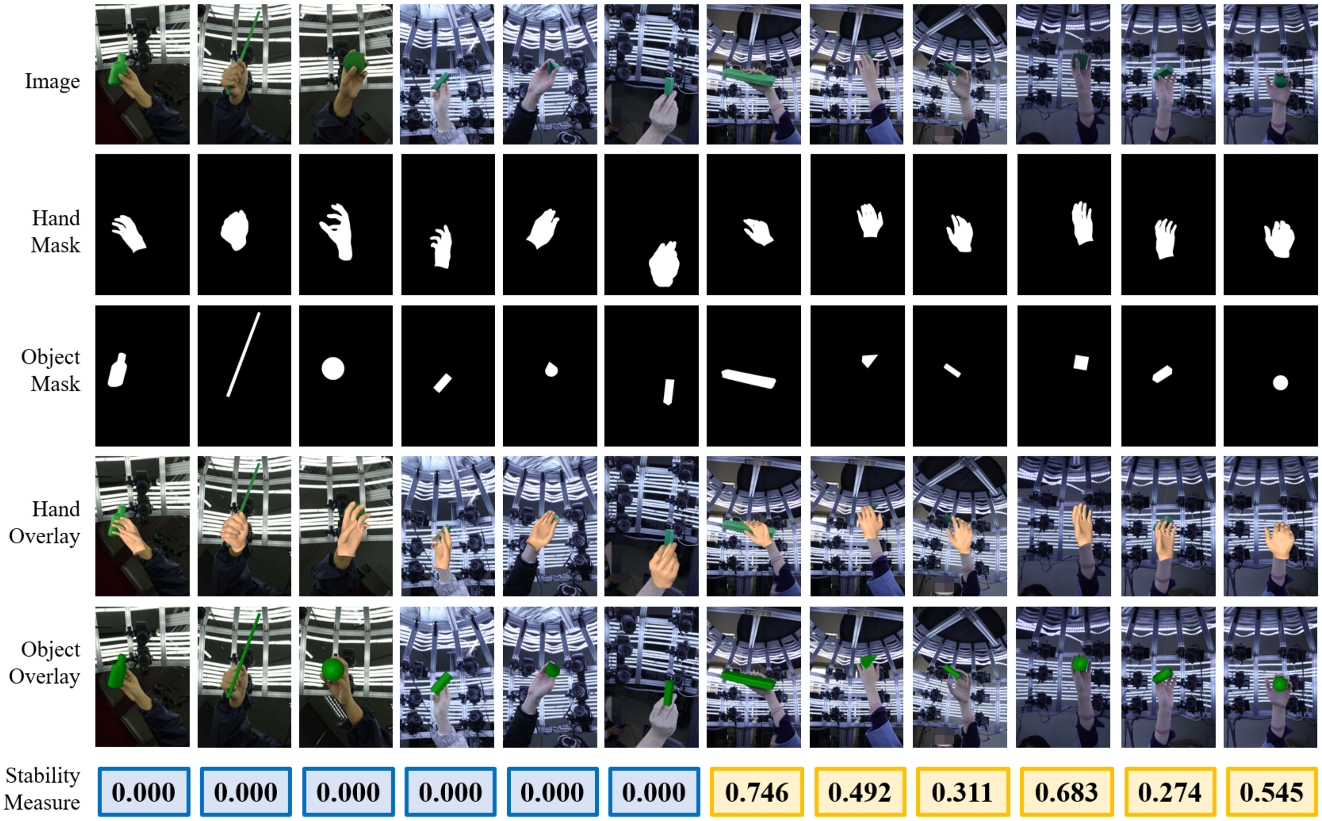


Figure 3. **Interaction dataset with balancing force measurements**. Each column corresponds to an interaction scene. The first 6 columns show the scene without threads, and the last 6 columns show the scene with the thread and balancing force on the thread. To facilitate the display, only one of the 25 perspective images is selected for display.

and the hand, the key-point order on the object is defined and marked in each perspective. By minimizing the reprojection error, this 2D information from multi-view images is used to optimize the shape and pose for the 3D object mesh and the hand LBS template.

Force Application. As shown in Fig. 2 (b), a dynamometer is hoisted on the top of the system. To flexibly measure forces from different directions, it is mounted on a tripod head with 3 DoF rotation. Due to the large difference in mass of all our objects, the dynamometers have different ranges and accuracies: (1) The range is 20N and the resolution is 0.1N; (2) The range is 10N and the resolution is 0.05N; (3) The range is 1N and the resolution is 0.02N. We chose thin threads made of UHMWPE (Ultra-high-molecular-weight polyethylene) material to connect the dynamometer and the interacted object. It is transparent and only 0.14mm in diameter. Some originally stable scenes without thin balancing thread are also captured. As shown in the first six columns and the last six columns of Fig. 3, this thin thread does not significantly affect the appearances of the interacted images. The capturing signal of the camera system was triggered when the indication of the dynamometer and the hand object interaction state is no longer changed. For each subject contacting each object, 5 actions with obvious differences in stability (indicated by the dynamometer) are collected. To facilitate postural fixation, the elbow participant was supported by the table plate during the capturing. To reduce the error, each scene is repeated 5 times through the above process, and the mean magnitude of the force is finally recorded.

C. More Alternatives

In this section, we provide evaluations of more variants with the same evaluating conditions as our main paper. These experiments focus on comparing the effects of different hyper-parameters in our sampling process on the final results. Among them, the number of samples N has the greatest impact on the results. Although a larger number of samples gives better results, for the hand model with 23 DoFs, we find that setting it to 300 is sufficient, and the improvement in results from more samples does not compensate for the longer running time of the algorithm. The number of sampling iterations K has a relatively small effect on the results. The number of the simulation time step T has a truncated effect on the method. When the step amount exceeds 150 steps, the results hardly change. We find

Method	Inter.(cm^3) ↓	Disp. (mm) ↓	SC. ↓
$K = 20$	6.28	1.15	0.36
$K = 40$	6.23	1.12	0.31
$N = 100$	7.03	1.13	0.39
$N = 500$	6.12	1.09	0.26
$T = 60$	6.35	1.31	0.41
$T = 240$	6.24	1.13	0.31
Ours	**6.24**	**1.13**	**0.31**

Table 2. **More Ablation study on ContactPose [3].** The components in network training paradigm, optimization function and physical hand model are evaluated.

that the final stable contact pattern obtained by our method is hardly affected when the physical parameters are varied within the same order of magnitude. Nevertheless, considering the simulation stability [1], the recommended values should be close to the parameters measured in our dataset.

References

[1] Pybullet. http://pybullet.org. 4

[2] Thingiverse. https://www.thingiverse.com/. 1

[3] Samarth Brahmbhatt, Chengcheng Tang, Christopher D Twigg, Charles C Kemp, and James Hays. Contactpose: A dataset of grasps with object contact and hand pose. In *ECCV*, pages 361–378. Springer, 2020. 4

[4] Bo Li, Lionel Heng, Kevin Koser, and Marc Pollefeys. A multiple-camera system calibration toolbox using a feature descriptor-based calibration pattern. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1301–1307. IEEE, 2013. 1

[5] Yangang Wang, Baowen Zhang, and Cong Peng. Srhandnet: Real-time 2d hand pose estimation with simultaneous region localization. *IEEE TIP*, 29:2977–2986, 2019. 2

Consent Form

Title of the study: Stability-driven Contact Reconstruction From Monocular Color Images

Purpose of the Study: The purpose of the study is to reconstruct the hand-object contact pattern directly from monocular images and utilize the physical stability criterion in the simulation to drive the optimization process.

Participation: I will cooperate with researchers to shoot the dataset. Before shooting the dataset, I will be asked the purpose of this task in detail. During the shooting process, I will also be given detailed guidance. I am interested in this research and agree to participate in the production of dataset.

Confidentiality and Privacy: I have received assurance from the researchers that the information I provided will remain strictly confidential. I understand that the dataset will be used only for academic purposes and that my identity will be protected.

Voluntary Participation: I am under no obligation to participate and if I choose to participate, I can withdraw from the study at any time and refuse to answer any questions, without suffering any negative consequences. If I choose to withdraw, everything related to me in the dataset will be removed and not used in the study. If I have any questions about the study, I can contact the researcher or their supervisor at any time.

It is recommended that I print a copy of this consent form for my records.

Acceptance: By signing my name below, I agree to participate in this research study.

龙治宇.	李治	刘星佑	乔佳奇	谢薇
祝忠钰	邱瀚宸	赵子萌	[illegible]	左炳辉
任奇牧	倪儒婷	于志鹏	方子祥	赵旭琦
王一白	潘水平	潘亮	顾杰	田[illegible]