SeqGrasp: Sequential Multi-Object Dexterous Grasp Generation

Haofei Lu, Yifei Dong, Zehang Weng, Jens Lundell, Florian Pokorny, and Danica Kragic

***Abstract*— We introduce the sequential multi-object robotic grasp sampling algorithm SeqGrasp that can robustly synthesize stable grasps on diverse objects using the robotic hand's partial Degrees of Freedom (DoF). We experimentally evaluate SeqGrasp against the state-of-the-art non-sequential multi-object grasp generation method MultiGrasp in simulation and on a real robot. The experimental results demonstrate that SeqGrasp reach an 8.71%-43.33% higher grasp success rate than MultiGrasp.**

I. Introduction

Generation of dexterous grasps has been studied for a long time, both from a technical perspective on generating grasps on robots [1]–[11] and understanding human grasping [12]–[15]. Most of these methods rely on bringing the robotic hand close to the object and then simultaneously enveloping it with all fingers. While this strategy often results in efficient and successful grasp generation, it simplifies dexterous grasping to resemble parallel-jaw grasping, thereby underutilizing the many DoF of multi-fingered robotic hands [10]. In contrast, grasping multiple objects with a robotic hand, particularly in a sequential manner that mirrors human-like dexterity, as shown in Fig. 1, is still an unsolved problem.

In this work, we introduce SeqGrasp, a novel hand-agnostic algorithm for generating sequential multi-object grasps. Our approach utilizes an optimization-based method to sequentially determine single-object grasp poses using a subset of the hand's DoF. As the grasp sequence progresses, the DoF engaged in previous grasps are frozen, leaving only the remaining DoF available for subsequent object grasps. To only engage a subset of the hand's DoF for each grasp, we propose an Opposition Space (OS) selection strategy that enables stable grasping using only a pair of links.

We experimentally evaluate SeqGrasp, and the state-of-the-art simultaneous multi-object grasping method MultiGrasp [16] in simulation and on physical hardware. The simulation results revealed that SeqGrasp and SeqDiffuser perform on par with MultiGrasp for picking one or two objects while outperforming it when picking three to four objects.

Our contributions can be summarized as follows:

- SeqGrasp, a novel hand-agnostic algorithm for sequential multi-object grasp generation.
- Extensive simulation experiments demonstrating the performance of SeqGrasp.

All authors are with the division of Robotics, Perception, and Learning (RPL) at KTH, Stockholm, Sweden. {haofeil,yifeid,zehang,jelundel,fpokorny, dani}@kth.se

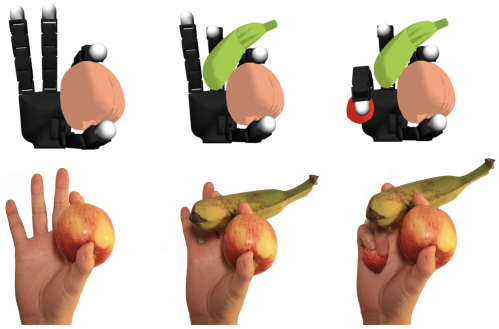


Fig. 1: **Sequential multi-object grasping.**

II. Related Work

A. Analytical Dexterous Grasping

Early research in dexterous grasping generated stable grasps by optimizing a grasp quality metric such as the force-closure metric [17], [18]. Although these methods are theoretically sound, they are computationally demanding because (i) the many DoF dexterous hands cause high-dimensional search spaces [19], and (ii) the quality metrics are expensive to compute [20]. Consequently, some methods have focused on reducing the search space by imposing constraints on the hand [21], [22] or restricting it to joint configuration subspaces [19]. Another line of work has proposed a computationally cheap differentiable force-closure estimator [10], [20], which has the advantage of being hand-agnostic. This work extends the differentiable force-closure measure from [10] to sequential multi-object grasping.

B. Multi-Object Grasping

Multi-object grasping presents unique challenges due to the complex multi-object interactions and the high-dimensional configuration space spanned by the hand and the objects. A few works have addressed dexterous multi-object grasping [16], [23], [24] where [16] targets simultaneous multi-object grasping while [23], [24] targets sequential multi-object grasping. Li et al. [16] proposed MultiGrasp a two-stage simultaneous multi-object dexterous grasping framework where a generative grasp sampler proposed poses to simultaneously pick many objects, followed by a learned policy for executing the pick. The main limitation of [16] is that objects must be spatially close and of similar size and shape. In comparison, our method can handle scattered objects of different shapes and sizes by sequentially picking one at a time. The other works that do sequential multi-object grasping [23], [24] restrict the grasping to a maximum of two objects [23] or to primitive object shapes such as cylinders

or spheres [24]. In comparison, our method can handle up to four objects of complex shapes and sizes.

III. Problem Formulation

The problem addressed in this work is sequential multi-object grasping, which we define as follows:

Definition 1 (Sequential multi-object grasping)**.** A sequential multi-object grasp is a grasp where one object is grasped at a time using a subset of the dexterous hand's DoF, while previously grasped objects, if any, remain fixed to the hand.

To contrast, *simultaneous multi-object grasping* addresses how to grasp multiple objects *simultaneously*, typically utilizing all the DoF of the hand [16].

We formulate the sequential multi-object grasping problem as generating a sequence of N grasps $\mathcal{G} = \{\mathbf{g}_i\}_{i=1}^{N}$ for picking a sequence of N objects $\mathcal{O} = \{\mathbf{O}_i\}_{i=1}^{N}$, where each $\mathbf{g}_n \in \mathcal{G}$ is restricted to a specific subset $\mathcal{OS}_n$ of the hand's total DoF and $N \geq 2$. Mathematically, this can be described as

$$\mathbf{g}_n = \underset{\mathbf{g}_n}{\operatorname{argmin}}\, E(\mathbf{g}_n, \mathbf{O}_n, \mathcal{G}_{n-1}, \mathcal{O}_{n-1}, \mathcal{OS}_n), \quad \forall n = 1, \ldots, N, \tag{1}$$

where $\mathcal{G}_{n-1} = \{\mathbf{g}_i\}_{i=1}^{n-1}$, $\mathcal{O}_{n-1} = \{\mathbf{O}_i\}_{i=1}^{n-1}$, $\mathcal{G}_0 = \varnothing$, and $\mathcal{O}_0 = \varnothing$. E in Eq. 1 is a differentiable function that quantifies how well grasp $\mathbf{g}_n$ can pick object $\mathbf{O}_n$ with the DoF $\mathcal{OS}_n$ given all previously generated grasps $\mathcal{G}_{n-1}$ and objects $\mathcal{O}_{n-1}$.

In this work, we represent $\mathcal{OS}_n$ as an opposition space (Section IV-A), each object $\mathbf{O} \in \mathcal{O}$ as a triangular mesh, and each grasp $\mathbf{g} \in \mathcal{G}$ as a vector $\mathbf{g} = [\mathbf{p}, \mathbf{r}, \boldsymbol{\theta}] \in \mathbb{R}^{9+K}$, where $\mathbf{p} \in \mathbb{R}^3$ is the hand's base position, $\mathbf{r} \in \mathbb{R}^6$ is the hand's base orientation in a 6D continuous reprsentation [25], and $\boldsymbol{\theta} \in \mathbb{R}^K$ is the K-dimensional hand joint angles which are 16 for the Allegro Hand. We assume the shape of all objects in $\mathcal{O}$ to be fully known. Next, we will introduce SeqGrasp our algorithm for solving Eq. 1.

IV. Sequential Grasp Generation

Here, we present Algorithm 1 for sequential grasp generation. It includes (i) an opposition space selection strategy (Section IV-A), (ii) an optimization-based grasp synthesis method (Section IV-B), and (iii) an energy-based cost function (Section IV-C).

A. Opposition Space Selection Strategy

The primary objective in sequential multi-object grasping is to maximize the hand's remaining DoF after each grasp. For this purpose, we propose a grasp planning strategy guided by OSes [14], [24], [26], [27]. An OS is a functional subspace within the hand's kinematic structure formed by pairs of opposing surfaces (such as fingertips, lateral surfaces of fingers, or palm surfaces) along with the joints that control these surfaces [24]. It represents regions where opposing forces can be applied to create stable grasps. The number of OSes is hand-dependent and varies based on the kinematic structure. Fig. 2a shows the seven different OSes for the Allegro Hand.

Algorithm 1: SeqGrasp

Input : Object sequence $\mathcal{O}$, OSes $\mathcal{OS}$, N_{step}, and p_{accept}.
Output: The optimized grasp sequence ${\mathcal{G}^*}_n$.

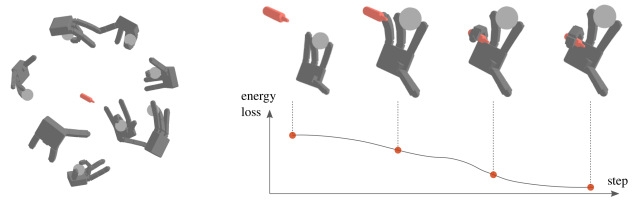
1 $n = 1$;
2 **while** $\mathcal{OS} \neq \varnothing$ *and* $n \leq N$ **do**
3 $\quad \mathcal{OS}_n \sim \mathcal{U}(\mathcal{OS})$;
4 $\quad \{\mathbf{x}_j\}_{j=1}^{2} \sim \mathcal{U}(\mathbf{S}_n)$;
5 $\quad$ **for** $s = 1$ **to** N_{step} **do**
6 $\quad\quad \Delta = \partial E(\mathbf{g}_n, \mathbf{O}_n, \mathcal{G}_{n-1}, \mathcal{O}_{n-1}, \{\mathbf{x}_j\}_{j=1}^{2})/\partial \mathbf{g}_n$;
7 $\quad\quad \mathbf{g}_n \leftarrow \text{MALA}(\mathbf{g}_n, \mathbf{J}_n, \Delta)$;
8 $\quad\quad \{\mathbf{x}_j\}_{j=1}^{2} \sim f(\mathbf{S}_n, p_{\text{accept}})$;
9 $\quad$ **end**
10 $\quad \mathcal{OS} \leftarrow \mathcal{OS} \setminus \mathcal{OS}_n$;
11 $\quad$ **for** $\mathcal{OS}_j \in \mathcal{OS}$ **do**
12 $\quad\quad \mathbf{J}_j \leftarrow \mathbf{J}_j \odot (\mathbf{1} - \mathbf{J}_n)$;
13 $\quad\quad$ **if** $\mathbf{J}_j = \mathbf{0}$ **then**
14 $\quad\quad\quad \mathcal{OS} \leftarrow \mathcal{OS} \setminus \mathcal{OS}_j$;
15 $\quad\quad$ **end**
16 $\quad$ **end**
17 $\quad n \mathrel{+}= 1$;
18 **end**

Mathematically, each opposition space can be represented as a pair $\mathcal{OS}_i = \{\mathbf{J}_i, \mathbf{S}_i\}$, where $\mathbf{J}_i \in \{0, 1\}^K$ is a binary vector indicating which joints are involved in controlling the opposition space, and $\mathbf{S}_i \in \mathbb{R}^{3 \times M_i}$ represents the 3D points on the hand where opposing forces can be applied. Fig. 2b shown an example of two different $\mathbf{S}_i$ for the Allegro Hand, where palm and pad oppositions have contact points located on the inner surfaces of fingers and palm and side oppositions have contact points on the fingers' lateral surfaces.

Let $\mathcal{OS} = \{\mathcal{OS}_i\}_{i=1}^{L}$ be the set of all OSes. Given this set, Algorithm 1 samples a random OS from it (Line 3) and uses it for subsequent grasp generation (Section IV-B). Once grasp generation is complete, the sampled OS can no longer be used and is thus removed from the available OSes (Line 10). $\mathbf{J}_i$ of all the remaining OSes are also updated by zeroing out the joints used in $\mathcal{OS}_n$ (Line 12). Subsequently, all OSes with $\mathbf{J} = \mathbf{0}$, meaning that no more controllable joints exist, are removed (Line 14). For instance, in the case of the

(a) (b)

Fig. 2: (a) **Grasps using all seven OSes**. From left to right, first row: middle-ring, index-middle, and thumb-index, second row: ring-palm, middle-palm, index-palm, and thumb-palm. (b) **Visualization of contact point candidates** on Allegro Hand surface. Cyan and pink points denote palm opposition and side opposition contacts, respectively.



(a) (b)

Fig. 3: (a) **Initialization.** The initial grasp configurations are randomly sampled on the expanded convex hull of the object (bottle) while the previously grasped object (ball) remains in the hand. (b) **Optimization.** During optimization, the grasp is incrementally refined, ultimately securing the target object using the ring-palm OS.

Allegro Hand, if the thumb-index OS is selected, then both the thumb-palm and index-palm OSes become unavailable due to shared joint constraints.

B. Optimization-based Grasp Generation

The next step in the algorithm (Lines 5-9) is to generate a stable, physically plausible, and collision-free grasp that respects the sampled $\mathcal{OS}_n$. To achieve this, we formulate E in Eq. 1 as an energy function (Section IV-C) and numerically optimize it using the Metropolis-Adjusted Langevin Algorithm (MALA) [28] (Line 7).

In robotic grasping, MALA has been used to optimize single object grasps $\mathbf{g}_n$ by iteratively refining $\mathbf{p}_n$, $\mathbf{r}_n$ and $\boldsymbol{\theta}_n$ according to Langevin dynamics [10], [20]. However, we must adapt MALA to sequential multi-object grasping. To this end, we propose a new grasp as $\hat{\mathbf{g}}_n \leftarrow \mathbf{g}_n - \gamma\,[\mathbf{1}, \mathbf{J}_n] \odot \Delta$, where γ is the step size, $\mathbf{1} \in \mathbb{R}^9$ is a padding vector to align the length of $\mathbf{J}$ with $\mathbf{g}$, $\Delta = \partial E / \partial \mathbf{g}_n$ is the energy gradient, and $\odot$ is the element-wise (Hadamard) product. $\hat{\mathbf{g}}_n$ is accepted if $\alpha \geq u$, where $u \sim \mathcal{U}([0,1])$ and

$$\alpha = \frac{E(\hat{\mathbf{g}}_n, \mathbf{O}_n, \mathcal{G}_{n-1}, \mathcal{O}_{n-1}, \{\mathbf{x}_j\}_{j=1}^2)}{E(\mathbf{g}_n, \mathbf{O}_n, \mathcal{G}_{n-1}, \mathcal{O}_{n-1}, \{\mathbf{x}_j\}_{j=1}^2)}. \tag{2}$$

The above procedure is repeated for a fixed number of steps where, at each step, $\{\mathbf{x}_j\}_{j=1}^2$ is re-sampled with probability p_{accept} (Line 8). This resampling process helps accelerate convergence and escape from local minimas [10], [20].

We initialize $\mathbf{g}_n$ at a randomly sampled position on the expanded convex hull of the target object $\mathbf{O}_n$ as exemplified in Fig. 3a,. If $n = 1$, then $\boldsymbol{\theta}_1$ is set to a natural open-hand and collision-free posture, while for $n \geq 2$, $\boldsymbol{\theta}_n = \boldsymbol{\theta}_{n-1}$. A visual example of the optimization process when grasping a second object is shown in Fig. 3b.

C. Energy Function

Numerically optimizing the energy function in Eq. 1 should result in stable, collision-free, joint-respecting, and OS-respecting grasps. We design the following energy function to capture all of these behaviors

$$E = \mathbf{w}^T \begin{bmatrix} E_{\text{fc}} & E_{\text{dis}} & E_{\text{hop}} & E_{\text{hsp}} & E_{\text{joint}} & E_{\text{oop}} \end{bmatrix}^T, \tag{3}$$

where $\mathbf{w} \in \mathbb{R}^6$ is a weight vector controlling the relative importance of the force-closure E_{fc}, contact distance E_{dis}, hand-object penetration E_{hop}, hand self-penetration E_{hsp}, joint limits E_{joint}, and object-object penetration E_{oop} energy terms.

The force-closure term (E_{fc}) encourages the grasp to be in force-closure equilibrium [29]. Following [10] and assuming zero friction and uniform contact force magnitudes, we define it as

$$E_{\text{fc}}(\{\mathbf{x}_j\}_{j=1}^2) = \|\mathbf{G}\mathbf{c}\|^2, \tag{4}$$

where $\mathbf{c} = [\mathbf{c}_1^T, \mathbf{c}_2^T]^T \in \mathbb{R}^{6\times 1}$ represents the concatenated contact normals at each contact point $\{\mathbf{x}_j\}_{j=1}^2$. $\mathbf{G}$ is defined as:

$$\mathbf{G} = \begin{bmatrix} \mathbf{I} & \mathbf{I} \\ [\mathbf{x}_1]_\times & [\mathbf{x}_2]_\times \end{bmatrix}, \tag{5}$$

where $\mathbf{I}$ represents the identity matrix, and $[\mathbf{x}_j]_\times$ $(1 \leq j \leq 2)$ denotes the skew-symmetric matrix formed from the contact point $\mathbf{x}_j$.

The contact distance and penetration terms (E_{dis} & E_{hop}) encourage the hand-object contacts to occur close to the object surface without penetrating it. The contact distance is mathematically defined as

$$E_{\text{dis}}(\{\mathbf{x}_j\}_{j=1}^2, \mathbf{O}_n) = \sum_{j=1}^{2} d(\mathbf{x}_j, \mathbf{O}_n), \tag{6}$$

where $d(\mathbf{x}_j, \mathbf{O}_n) = \min_{\mathbf{v}\in\mathbf{O}_n} \|\mathbf{x}_j - \mathbf{v}\|_2$ is the shortest point-mesh distance. Similarly, the hand-object penetration term is defined as:

$$E_{\text{hop}}(\mathbf{g}_n, \mathbf{O}_n) = \sum_{\mathbf{v}\in\mathcal{V}_{\text{hop}}(\mathbf{H}_\mathbf{g}, \mathbf{O}_n)} d(\mathbf{v}, \mathbf{O}_n), \tag{7}$$

where $d(\mathbf{v}, \mathbf{O}_n) = \min_{\mathbf{v}_1\in\mathbf{O}_n} \|\mathbf{v} - \mathbf{v}_1\|_2$ and $\mathcal{V}_{\text{hop}}(\mathbf{H}_\mathbf{g}, \mathbf{O}_n)$ is the set of points on the hand surface pointcloud $\mathbf{H}_\mathbf{g} \in \mathbb{R}^{3\times M_h}$ that penetrate the object $\mathbf{O}_n$.

The self-collision and joint limit terms (E_{hsp} & E_{joint}) encourage physical feasibility. We define these as

$$E_{\text{hsp}}(\mathbf{g}_n) = \sum_{\mathbf{v}_1, \mathbf{v}_2 \in \mathcal{V}_{\text{hsp}}(\mathbf{H}_\mathbf{g}), \mathbf{v}_1 \neq \mathbf{v}_2} \max(\|\mathbf{v}_1 - \mathbf{v}_2\|_2, 0), \tag{8}$$

$$E_{\text{joint}}(\mathbf{g}_n) = \|(\boldsymbol{\theta} - \boldsymbol{\theta}^{\text{upper}})^+\|_1 + \|(\boldsymbol{\theta}^{\text{lower}} - \boldsymbol{\theta})^+\|_1, \tag{9}$$

where $\mathcal{V}_{\text{hsp}}(\mathbf{H}_\mathbf{g})$ denotes all surface points of the hand that are self-penetrating, $(\cdot)^+$ denotes the element-wise operation $\max(\cdot, 0)$, and $\boldsymbol{\theta}^{\text{upper}}$ and $\boldsymbol{\theta}^{\text{lower}}$ denote the upper and lower limits of all joints.

Finally, the term (E_{oop}) minimizes object-object penetration. It is defined as

$$E_{\text{oop}}(\{\mathbf{O}_i\}_{i=1}^n) = \sum_{i=1}^{n-1} \sum_{\mathbf{v}\in\mathcal{V}_{\text{oop}}(\mathbf{O}_i, \mathbf{O}_n)} d(\mathbf{v}, \mathbf{O}_n), \tag{10}$$

where $d(\mathbf{v}, \mathbf{O}_n) = \min_{\mathbf{v}_1\in\mathbf{O}_n} \|\mathbf{v} - \mathbf{v}_1\|_2$, and $\mathcal{V}_{\text{oop}}(\mathbf{O}_i, \mathbf{O}_n)$ are the inter-penetrating surface points between the previously grasped object $\mathbf{O}_i$ and the current object $\mathbf{O}_n$.

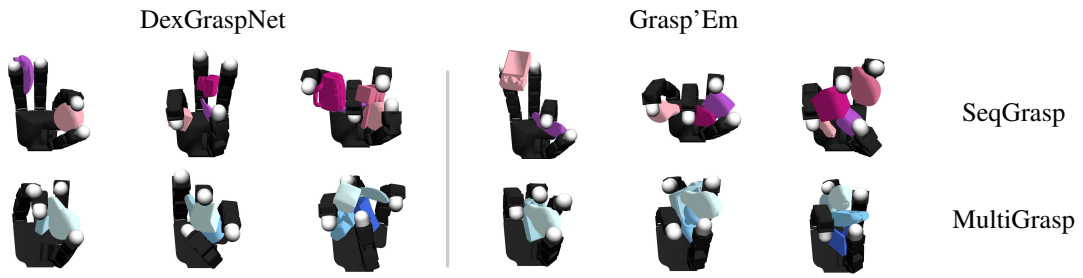


Fig. 4: **Qualitative results.** For SeqGrasp, we only show consumed grasps, that is, when $\mathcal{OS} = \emptyset$. For SeqGrasp, the grasp sequences are visually indicated by a color gradient, transitioning from lighter to darker shades. In contrast, for MultiGrasp, the color gradient is only used to differentiate the objects.

V. Experiments

We compare our method to the optimization-based sampler MultiGrasp from [16], which generates simultaneous grasps on clustered objects. As such, for MultiGrasp, we must first sample clustered object configurations and then generate multi-object grasps directly on the object cluster. In contrast, SeqGrasp do not require objects to be spatially close as they generate grasps sequentially based on previously successful ones. While this comparison is not entirely fair, we still believe comparing these two strategies offers valuable insights.

We evaluated all generated grasps in the simulation experiments in Isaac Gym [30]. We used two object sets: (1) all eight objects from Grasp'Em [16] and (2) a random selection of eight validation objects from DexGraspNet [10] object set. We randomly generated ten four-object sequences for each object set, and, per object, we generated 256 grasps, resulting in 10,240 grasps per method.

We used the following metrics to assess the quality of the generated grasps:

1) **Success rate (SR) in percent**: Following the experiment setup in [10], a grasp is considered as successful if all objects within the hand can resist an acceleration of 9.8 $\mathrm{m/s^2}$ in all six orthogonal directions for 100 consecutive simulation steps.
2) **Maximum penetration depth (Pene.) in mm**: The maximum interpenetration distance between the hand and all grasped objects.
3) **Diversity (Div.) in radian**: Grasp diversity is determined by calculating the standard deviation of $\mathbf{g}$ across all successful grasps.

	SR ↑		Pene. ↓		Div. ↑
Method	**SData**	**G'Em**	**SData**	**G'Em**	**Avg.**
MulG-1	**66.84**	**65.39**	**1.14**	**1.27**	0.284
SeqG-1	50.04	40.78	1.73	2.16	**0.332**
MulG-2	**22.46**	16.48	2.30	2.83	0.347
SeqG-2	21.21	**32.03**	**2.14**	**1.78**	**0.367**
MulG-3	10.78	3.55	3.39	4.04	0.340
SeqG-3	**19.49**	**21.05**	**2.23**	**2.23**	**0.349**
MulG-4	0.90	0.47	5.17	6.27	**0.329**
SeqG-4	2.93	**5.04**	**2.70**	**2.62**	0.312

TABLE I: **Simulation results**. MulG, SeqG, G'Em, and SData are short the MultiGrasp, SeqGrasp, Grasp'Em, and SeqDataset, respectively. The $-i$ following the method name denotes the number of objects used for grasp generation. ↑(↓), the higher (lower), the better.

The quantitative results are presented in Table I, while Fig. 4 qualitatively illustrates a few grasps. The results demonstrate that SeqGrasp achieves the higher success rate and the lower penetration depth when grasping two or more objects. MultiGrasp performs well for one- and two-object grasps, as it utilizes all of the hand's available DoF to grasp the objects. However, because MultiGrasp requires all objects to be initialized nearby, the success rate of the generated grasps is susceptible to the initial object placements. In contrast, SeqGrasp do not suffer from this limitation.

We observe a significant performance drop when transitioning from three-object to four-object grasps across all methods. We hypothesize that this decline occurs because the three previously grasped objects occupy substantial space within the Allegro Hand, pushing the fourth object grasp to the limits of the hand's kinematic redundancy. Additionally, as the number of grasped objects increases, object-object interactions grow exponentially, making the task considerably more challenging, a finding also reported in [16]. Nevertheless, SeqGrasp demonstrates superior performance in scenarios involving three or more objects.

VI. Conclusion

We proposed SeqGrasp, an algorithm for sequentially grasping multiple objects with a dexterous hand. SeqGrasp combines OSes and differentiable-force closure to generate stable grasps that maximize the hand's remaining DoF after each grasp. The experimental evaluations demonstrated that SeqGrasp outperformed the simultaneous multi-object grasping baseline MultiGrasp, achieving a higher average success rate when grasping three to four objects. In conclusion, this work demonstrated a stable sequential multi-object grasp generation solution, which we hope can pave the way for more research in multi-object grasping.

REFERENCES

[1] J. Bohg, A. Morales, T. Asfour, and D. Kragic, "Data-driven grasp synthesis—a survey," *IEEE Transactions on robotics*, vol. 30, no. 2, pp. 289–309, 2013.

[2] M. Li, K. Hang, D. Kragic, and A. Billard, "Dexterous grasping under shape uncertainty," *Robotics and Autonomous Systems*, vol. 75, pp. 352–364, 2016.

[3] A. Billard and D. Kragic, "Trends and challenges in robot manipulation," *Science*, vol. 364, no. 6446, p. eaat8414, 2019.

[4] R. Newbury, M. Gu, L. Chumbley, A. Mousavian, C. Eppner, J. Leitner, J. Bohg, A. Morales, T. Asfour, D. Kragic *et al.*, "Deep learning approaches to grasp synthesis: A review," *IEEE Transactions on Robotics*, vol. 39, no. 5, pp. 3994–4015, 2023.

[5] Z. Weng, H. Lu, D. Kragic, and J. Lundell, "Dexdiffuser: Generating dexterous grasps with diffusion models," *IEEE Robotics and Automation Letters*, vol. 9, no. 12, pp. 11 834–11 840, 2024.

[6] J. Lu, H. Kang, H. Li, B. Liu, Y. Yang, Q. Huang, and G. Hua, "Ugg: Unified generative grasping," in *European Conference on Computer Vision*. Springer, 2024, pp. 414–433.

[7] J. Lundell, F. Verdoja, and V. Kyrki, "Ddgc: Generative deep dexterous grasping in clutter," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 6899–6906, 2021.

[8] Z. Wei, Z. Xu, J. Guo, Y. Hou, C. Gao, Z. Cai, J. Luo, and L. Shao, "D(r,o) grasp: A unified representation of robot and object interaction for cross-embodiment dexterous grasping," *arXiv preprint arXiv:2410.01702*, 2024.

[9] Y. Xu, W. Wan, J. Zhang, H. Liu, Z. Shan, H. Shen, R. Wang, H. Geng, Y. Weng, J. Chen, T. Liu, L. Yi, and H. Wang, "Unidexgrasp: Universal robotic dexterous grasping via learning diverse proposal generation and goal-conditioned policy," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2023, pp. 4737–4746.

[10] R. Wang, J. Zhang, J. Chen, Y. Xu, P. Li, T. Liu, and H. Wang, "Dexgraspnet: A large-scale robotic dexterous grasp dataset for general objects based on simulation," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 11 359–11 366.

[11] P. Li, T. Liu, Y. Li, Y. Geng, Y. Zhu, Y. Yang, and S. Huang, "Gendexgrasp: Generalizable dexterous grasping," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 8068–8074.

[12] J. Romero, H. Kjellström, and D. Kragic, "Hands in action: real-time 3d reconstruction of hands in interaction with objects," in *2010 IEEE International Conference on Robotics and Automation*. IEEE, 2010, pp. 458–463.

[13] D. Song, N. Kyriazis, I. Oikonomidis, C. Papazov, A. Argyros, D. Burschka, and D. Kragic, "Predicting human intention in visual observations of hand/object interactions," in *2013 IEEE International Conference on Robotics and Automation*. IEEE, 2013, pp. 1608–1615.

[14] 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.

[15] M. Kokic, D. Kragic, and J. Bohg, "Learning task-oriented grasping from human activity datasets," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 3352–3359, 2020.

[16] Y. Li, B. Liu, Y. Geng, P. Li, Y. Yang, Y. Zhu, T. Liu, and S. Huang, "Grasp multiple objects with one hand," *IEEE Robotics and Automation Letters*, vol. 9, no. 5, pp. 4027–4034, 2024.

[17] A. Miller and P. Allen, "Graspit! a versatile simulator for robotic grasping," *IEEE Robotics & Automation Magazine*, vol. 11, no. 4, pp. 110–122, 2004.

[18] C. Rosales, R. Suárez, M. Gabiccini, and A. Bicchi, "On the synthesis of feasible and prehensile robotic grasps," in *IEEE International Conference on Robotics and Automation*. IEEE, 2012, pp. 550–556.

[19] M. T. Ciocarlie and P. K. Allen, "Hand posture subspaces for dexterous robotic grasping," *The International Journal of Robotics Research*, vol. 28, no. 7, pp. 851–867, 2009.

[20] T. Liu, Z. Liu, Z. Jiao, Y. Zhu, and S.-C. Zhu, "Synthesizing diverse and physically stable grasps with arbitrary hand structures using differentiable force closure estimator," *IEEE Robotics and Automation Letters*, vol. 7, no. 1, pp. 470–477, 2022.

[21] J. Ponce, S. Sullivan, J.-D. Boissonnat, and J.-P. Merlet, "On characterizing and computing three- and four-finger force-closure grasps of polyhedral objects," in *Proceedings IEEE International Conference on Robotics and Automation*, 1993, pp. 821–827 vol.2.

[22] J.-W. Li, H. Liu, and H.-G. Cai, "On computing three-finger force-closure grasps of 2-d and 3-d objects," *IEEE Transactions on Robotics and Automation*, vol. 19, no. 1, pp. 155–161, 2003.

[23] S. He, Z. Shangguan, K. Wang, Y. Gu, Y. Fu, Y. Fu, and D. Seita, "Sequential multi-object grasping with one dexterous hand," *arXiv preprint arXiv:2503.09078*, 2025.

[24] K. Yao and A. Billard, "Exploiting kinematic redundancy for robotic grasping of multiple objects," *IEEE Transactions on Robotics*, vol. 39, no. 3, pp. 1982–2002, 2023.

[25] Y. Zhou, C. Barnes, J. Lu, J. Yang, and H. Li, "On the continuity of rotation representations in neural networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 5745–5753.

[26] T. Iberall, "Opposition space as a structuring concept for the analysis of akilled hand movements," *Experimental brain research series*, vol. 15, pp. 158–173, 1986.

[27] J. B. Smeets, K. van der Kooij, and E. Brenner, "A review of grasping as the movements of digits in space," *Journal of neurophysiology*, vol. 122, no. 4, pp. 1578–1597, 2019.

[28] G. O. Roberts and R. L. Tweedie, "Exponential convergence of langevin distributions and their discrete approximations," *Bernoulli*, vol. 2, pp. 341–363, 1996.

[29] E. Rimon and J. Burdick, "On force and form closure for multiple finger grasps," in *Proceedings of IEEE International Conference on Robotics and Automation*, vol. 2. IEEE, 1996, pp. 1795–1800.

[30] V. Makoviychuk, L. Wawrzyniak, Y. Guo, M. Lu, K. Storey, M. Macklin, D. Hoeller, N. Rudin, A. Allshire, A. Handa *et al.*, "Isaac gym: High performance gpu-based physics simulation for robot learning," *arXiv preprint arXiv:2108.10470*, 2021.