

Grasping by interconnection: robust closing motions from coarse object templates

Julien Vanderheyden¹, Guillaume Drion¹, Fulvio Forni², Pierre Sacré¹

Abstract—Dexterous robot hands must often grasp objects whose shape, size, and pose are known only approximately. Grasp planners typically require accurate object models or correct errors with feedback, but how much inaccuracy a closing motion can tolerate on its own remains unclear. To address this question, we designed a motion planner based on four principles: a coarse template of the object, human grasp types, an object-centric interaction, and compliant, sliding contacts instead of prescribed contact points. This paper presents the planner, implemented through virtual model control, and its evaluation on a Shadow Dexterous Hand. Without feedback, the planned closing motions tolerated size errors of about 1 cm and pose errors of several centimeters and tens of degrees, a wider range than a state-of-the-art data-driven planner in 25 of 27 tested conditions. They also grasped 82.5 % of 80 everyday objects and succeeded within an autonomous pipeline. Robustness can thus be designed into the closing motion itself, rather than left only to feedback. This planner opens a path toward reliable manipulation in uncertain settings, which we will pursue by combining it with adaptive feedback control on the physical hand.

I. INTRODUCTION

Dexterous grasping in unstructured environments remains an open challenge in robotics [1]. The distributed, discontinuous contacts between the hand and the object often lead to failures, which planners mitigate with detailed information about the object alongside reliable, accurate sensing. In practice, however, object details are uncertain, and visual and tactile sensing are noisy. Reported pose estimation errors reach several centimeters in translation and several tens of degrees in orientation [2], the same magnitude as the spacing between the fingers. At this level of error, a planner cannot determine which part of the object each finger will touch.

Current approaches to grasping typically rely on accurate object models, obtained either from a library, which limits generalization, or from sensor-based reconstruction, which introduces its own challenges. Visual perception, for example, is often compromised by clutter, occlusions, and changing lighting, which lead to noisy or incomplete object estimates. Other approaches reduce the need for accurate models by relying on continuous corrections based on tactile or force feedback [3]. However, this makes it difficult to assess whether the planned motion was effective: is successful grasping the result of good motion planning or good

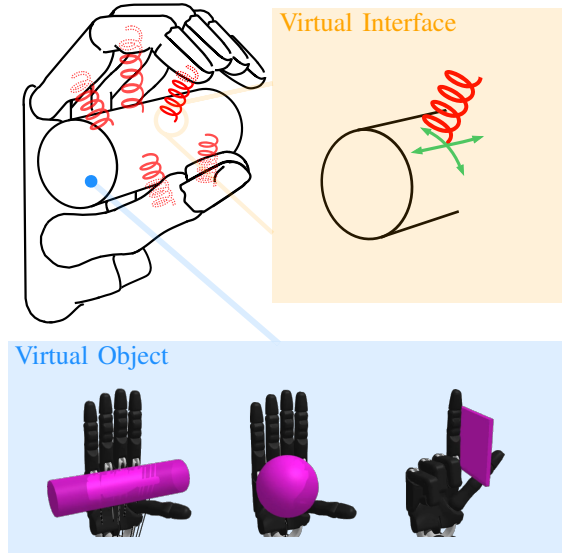


Fig. 1. Grasping by interconnection: the closing motion results from coupling the hand to a coarse template of the object, not from a prescribed trajectory. Each template (cylinder, sphere, or flat box) defines a *virtual object* (bottom). A *virtual interface* couples the virtual hand to this object through virtual springs and dampers (top); their terminals slide on the template surface (green arrows), so no contact point is prescribed. The motion is generated by virtual model control (VMC).

feedback adaptation and control? At the fundamental level, these approaches hide the robustness problem inherent in grasping: how much inaccuracy in the object description can a closing motion tolerate on its own, without additional sensing or feedback correction?

To answer this question, we propose a new motion planner that starts from a coarse description of the object to generate the closing motion of each finger. This planner is based on four principles. First, accurate object estimation is not realistic, so the planner starts from a *coarse template* of the object: a cylinder, a sphere, or a flat box, with its pose and size. Second, the motion of an anthropomorphic hand should take advantage of biological evolution, so the planner reuses *human grasps*: each template is paired with a grasp type from the human grasp taxonomy [4], [5], and the fingers coordinate following human motor behavior. Third, grasping should be object-centric, so the planner is based on an *object-centric interaction* between the hand and the object rather than assigning a target position to each finger. Fourth, compliance plays a major role in managing contacts, so the planner relies on simulated *compliant contacts*: contact points are not prescribed but derived from simulated physical interactions.

This work was supported by the Belgian Government through the Federal Public Service Policy and Support.

¹J. Vanderheyden, G. Drion, and P. Sacré are with the Department of Electrical Engineering and Computer Science, University of Liège, Belgium (julien.vanderheyden@uliege.be; gdrion@uliege.be; p.sacre@uliege.be).

²F. Forni is with the Department of Engineering, University of Cambridge, United Kingdom (f.forni@eng.cam.ac.uk).

The motion of each finger is then planned through a *virtual mechanical interface* between a virtual representation of the robot hand and the coarse templates. This approach is inspired by classical virtual model control (VMC) [6]–[9]. Coupling a hand to a template amounts to connecting mechanical elements, and passivity keeps the interaction stable without requiring exact contact models. This approach suits our principles: it is inherently object-centric and compliant, human grasp taxonomy is enforced through the mechanical structure of the virtual interface, and the contact points of the fingers are free to move on the surface of the object template, and are therefore not predefined. As illustrated in Fig. 1, the virtual interface generates virtual *forces* that drive the motion of the virtual fingers. The closing motion is thus planned by simulation. We call this approach *grasping by interconnection*. Once the planned motion is computed, the physical hand tracks it. No measurements of the real object are used during closure.

This paper presents the planner and the evidence for its robustness. Section II contextualizes the planner within the wider landscape of grasp planning. Section III describes how it implements the four principles. Section IV reports experiments on a Shadow Dexterous Hand [10], covering errors in object size and pose, variations of the stiffness and damping parameters, everyday objects, and a comparison with a state-of-the-art data-driven planner. The planned motions tolerated size errors of about 1 cm and pose errors of several centimeters and tens of degrees, a wider range than $\mathcal{D}(\mathcal{R}, \mathcal{O})$ Grasp [11] in 25 of 27 tested conditions, and grasped 82.5 % of 80 everyday objects.

The scope of this paper is deliberately narrow. We only address the planning of the closing motion of the hand, starting from a given pre-grasp pose. Reaching, clutter, and the selection of the object template are essential to any grasping pipeline but fall beyond this scope. Our planner does not estimate the description inaccuracy and does not adapt to it. It reads a template and generates one closing motion. We view this planner as a first step toward closed-loop grasping by interconnection.

II. RELATED WORK

A. Standard grasping approaches

Analytical and data-driven grasping methods achieve strong performance under controlled conditions, but depend on accurate object models, which limits their applicability in uncertain environments.

Analytical methods compute force-closure grasps by optimizing over contact geometry [12], [13]. These techniques inherently require detailed object models, either known *a priori*, which often limits their use to a predefined set of objects [14], or reconstructed online from partial observations, which introduces errors due to the noise of 3D sensing [15]. In both cases, the reliability of the planned grasp is limited by the accuracy of the object representation or reconstruction.

Data-driven approaches learn grasp strategies from large datasets, implicitly constructing internal object representations. They generalize to unseen objects either by recon-

structing the object shape [16] or by inferring grasps directly from point clouds [11]. Both strategies assume that the input data at test time matches the training distribution, which makes them sensitive to sensor noise and domain shift.

Beyond their reliance on accurate object models, both approaches typically output a final grasp, defined by contact points or a grasp pose, and leave the motion toward it to a generic planner or controller. When the size or pose of the object is wrong, contacts occur before or away from the predicted points, and the grasp outcome depends on a motion that was not designed to handle these errors. Our planner mitigates these issues through an object-centric closing motion relative to a coarse template. In contrast to predefined contact points, fingers track a coordinated closing motion adapted to the broad shape of the object.

B. Object shape approximation

An alternative to precise object reconstruction is to build approximate models and trade geometric detail for robustness [17]. Cognitive science supports this choice: human object recognition and manipulation favor simplified geometric approximations [18], [19].

Among the simplest strategies, shape primitives, such as cylinders, spheres, and boxes [17], [20], and box-based representations [21]–[23] have been successful in finding grasp regions. Richer representations, such as superquadrics [24], [25], medial axes [26], [27], and skeletal models [28], capture more geometric detail. However, their dependence on accurate shape estimation makes them more vulnerable to sensor noise. This trade-off between geometric expressiveness and robustness motivates the use of *coarse templates*.

Our planner differs from the methods above in how the template is used. These methods derive a target hand posture, leaving the associated motion planning to external algorithms, such as the robot hand controller. In contrast, we use templates to determine the entire closing motion. As shown in Figs. 1 and 2, each coarse template is paired to a specific virtual mechanism whose terminals slide on the template surface. The planned motion of each finger is then derived by simulation. This *object-centric interaction* connects to the evidence from cognitive science, adopting a coarse object geometry to shape the motion.

C. Human-inspired grasp synthesis

Human grasping has evolved to remain robust to uncertainty in perception, object geometry, and physical interaction, making it a natural source of inspiration for robust manipulation. It also provides a principled way to manage the high dimensionality of the grasping problem.

On the one hand, *grasp taxonomies* organize grasps into canonical types, providing lightweight priors that guide hand configuration without a full high-dimensional optimization. Explicit models of grasp type improve task success, even in data-driven settings [29]. However, early approaches paired each grasp type with a fixed closing motion [20], [30]–[32], which cannot adapt to geometric variations between objects. Methods that combine grasp types with analytical

optimization [28], [33]–[37] still require a detailed object model to evaluate grasp quality, and methods that use grasp types in learned policies [38]–[41] inherit the sensitivity to sensor noise discussed in Section II-A. The benefits of grasp taxonomies are thus often hindered by rigid or model-dependent executions.

On the other hand, *hand synergies* reduce the complexity of hand control by exploiting the low-dimensional structure of natural hand motion. Postural synergies [42], [43] capture the dominant modes of joint covariation in human grasping and provide a compact basis of hand postures that makes optimization and learned control tractable in real time [38], [44]. However, the synergy basis is extracted from hand postures rather than from the objects being grasped: the same basis applies regardless of object shape, which leaves open how synergies cope with errors in the object description.

In our approach, we bridge these two paradigms. We take advantage of grasp types to identify high-quality grasp configurations, which are later adapted to the variations in object size and geometry. Rather than constraining the hand to a predefined synergy subspace, we use VMC to coordinate the fingers through an *object-centric interaction*. A structured coupling of virtual springs and dampers between the hand and the template compliantly constrains the motion of the hand, without limiting its degrees of freedom. The taxonomy determines which coupling is appropriate for a given template, and VMC handles finger coordination. This design aims to retain the robustness of human grasping, as verified in Section IV.

III. GRASPING BY INTERCONNECTION

This section describes what our motion planner reads, what it computes, and how it implements the four principles. Section III-A specifies the coarse template and its grasp type. Section III-B explains why virtual model control implements an object-centric interaction and how the virtual hand is simulated. Sections III-C and III-D detail the two components of the planner: a virtual object, which models the template with compliant elements, and a virtual interface, which couples the virtual hand to it through sliding terminals and couplings derived from human grasps (Fig. 2).

A. Object description

Objects are reduced to *coarse templates*. The planner relies on (i) the template—a cylinder, a sphere, or a flat box—and (ii) a small set of parameters related to the pose and the size of the template. The position and, for cylinders and flat boxes, the orientation of the template set the pre-grasp pose of the hand. The closing motion itself depends only on the template size: the diameter of a cylinder or a sphere, or the thickness and width of a flat box. A template therefore consists of a discrete class with at most eight parameters, of which at most two shape the closing motion.

The mechanical interconnection between the template and the virtual hand encodes the *human grasp taxonomy*: it pairs the cylinder, the sphere, and the flat box with the medium wrap, the power sphere, and the lateral pinch,

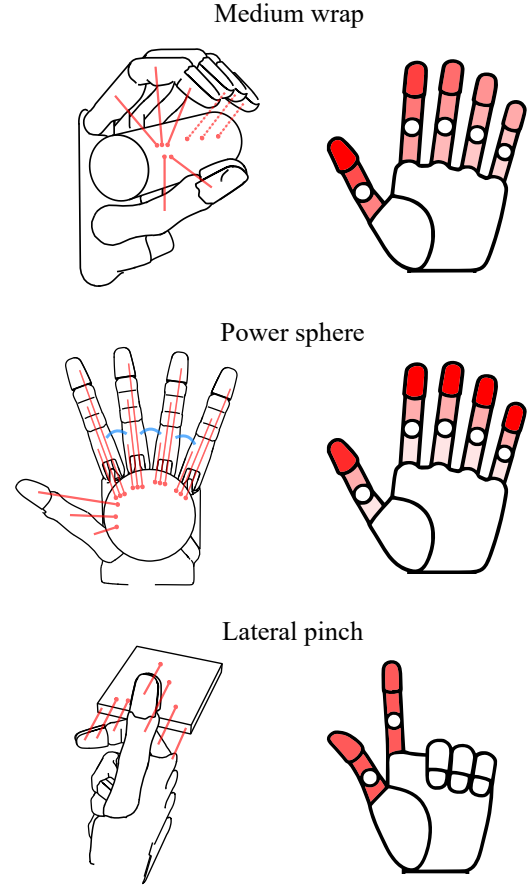


Fig. 2. The virtual interface encodes each grasp type through its architecture and stiffness distribution, without prescribing contact points. *Left*. Contractive (red) and repulsive (blue) springs drive the fingers toward stable contact; sliding terminals (circles) move on the template surface, so contact locations result from the motion. *Right*. Stiffness distribution set by α and β ; higher color intensity indicates higher stiffness.

respectively [4]. Bullock et al. [5] identified these three grasp types as the most versatile ones in household settings. Within each grasp type, the fingers’ motion is coordinated following human motor behavior, as detailed in Section III-D.

No further visual, tactile, or force measurements are used. Following the *compliant contacts* principle, the planner prescribes no contact point: contact locations are an outcome of the simulation of the virtual hand coupled to a template-based virtual mechanism, as described below. The tolerance reported in Section IV is therefore a property of the planned motion, executed without feedback from the object.

B. Virtual model control

Inspired by VMC [7], [8], our planner frames grasping as an *object-centric compliant interaction* realized by a mechanical interconnection between the virtual hand and the template object, rather than as a set of finger contact points. The associated (simulated) forces determine the planned coordinated motion of the fingers. In the classical case of robot manipulators with revolute joints, these virtual forces F_i act along coordinates $z_i = h_{z_i}(q)$ of the virtual elements,

such as the extension of a spring, expressed as functions of the joint coordinates $\mathbf{q}$. These forces are translated into joint torques by

$$\mathbf{u} = \sum_{i=1}^n J_{\mathbf{z}_i}^T(\mathbf{q}) \mathbf{F}_i, \quad (1)$$

where $J_{\mathbf{z}_i} = \frac{\partial h_{\mathbf{z}_i}(\mathbf{q})}{\partial \mathbf{q}}$ is the corresponding Jacobian matrix.

VMC guarantees that the controlled robot remains passive [45]–[48], within the limits of faithful approximation of the virtual mechanical elements. Passivity ensures stable interaction with passive objects without precise contact models, even at stiff contact transitions. Moreover, the use of virtual mechanism is inherently modular: grasp-specific elements can be added or reconfigured independently, which translates into highly configurable virtual interfaces for each template and grasp type. Finally, deriving finger motion via VMC and simulation does not require inverse kinematics computation, reducing the computational burden that typically arises from the high dimensionality of anthropomorphic hands.

We emphasize that, in this paper, we use VMC as a motion planner rather than as a feedback controller, so that the closing motion is fixed before execution and its robustness can be assessed on its own. The virtual hand is a copy of the robot hand built from virtual links, joints, masses, and inertias. The virtual object and the virtual interface described below, which together form the virtual mechanism, are attached to it, and the resulting mechanical system is simulated. The simulation integrates the dynamics of this virtual system,

$$M(\mathbf{x})\ddot{\mathbf{x}} + C(\mathbf{x}, \dot{\mathbf{x}})\dot{\mathbf{x}} + g(\mathbf{x}) = \sum_{i=1}^n J_{\mathbf{z}_i}^T(\mathbf{x}) \mathbf{F}_i, \quad (2)$$

where $\mathbf{x} = (\mathbf{q}, \mathbf{q}_v)$ collects the joint coordinates $\mathbf{q}$ of the virtual hand and the coordinates $\mathbf{q}_v$ of the virtual mechanism, namely the revolute and prismatic joints of the sliding terminals, which have low-inertia endpoints. M , C , and g are the inertia matrix, the Coriolis terms, and the gravity terms of this coupled system, and the Jacobians are taken with respect to $\mathbf{x}$. The output of the planner is the joint trajectory $\mathbf{q}(t)$ produced by this simulation, which the physical hand then tracks as a position reference. The simulation runs ahead of execution.

C. Virtual object

Each coarse template is a virtual object with a *compliant surface*, whose mechanics is regulated by simple springs and dampers. The surface springs generate repulsive forces that keep the virtual fingers from penetrating the virtual object. The force of each virtual spring is ReLU shaped

$$\mathbf{F}_s = k_{\text{obj}} \min(|\mathbf{z}| - z_{\text{rest}}, 0) \frac{\mathbf{z}}{|\mathbf{z}|}, \quad (3)$$

where $\mathbf{z}$, z_{rest} , and k_{obj} are the extension, rest length, and stiffness of the spring, respectively. Each spring pushes back only when compressed below its rest length, which mimics the onset of contact.

The template size sets the geometry of the virtual object and the mechanical parameters of the virtual interface: the rest lengths z_{rest} of the repulsive springs and the placement of the sliding terminals, so that the attractive and repulsive forces drive the fingers to the surface. The stiffness and damping parameters do not depend on the size.

The dampers suppress oscillations at contact. The associated forces satisfy $\mathbf{F}_d = c_{\text{obj}}(|\mathbf{z}|) \dot{\mathbf{z}}$ with ReLU-shaped damping

$$c_{\text{obj}}(|\mathbf{z}|) = \frac{c_{\text{max}}}{\gamma z_{\text{rest}}} \max(\gamma z_{\text{rest}} - |\mathbf{z}|, 0). \quad (4)$$

The coefficient $\gamma = 1.05$ inflates the virtual object by 5%, so that damping activates before the spring force. All force profiles are continuous, which avoids abrupt variations in the planned motion. Other passive contact models could replace these springs and dampers without changing the architecture.

D. Virtual interface

The virtual interface couples the virtual hand to the virtual object. Grasp synthesis combines two objectives: the identification of mechanically stable final contact configurations and the derivation of the motion trajectories that bring the hand to that configuration. In addition, different grasp types impose specific constraints that must be explicitly encoded. The three virtual interfaces of Fig. 2 address all these aspects.

Stable contact configurations follow from compliant interaction: the terminals of the contractive springs within the virtual interface are free to move on the virtual object surface. Therefore, they attract fingers and palm toward the template without forcing predefined contact points. These terminals slide along the template surface through prismatic and revolute joints, allowing for an entire set of final configurations, rather than a single one.

The transient motion of the fingers toward the final, stable configuration is shaped by the relative stiffness and damping of the virtual elements across the whole hand. Two complementary mechanisms, both inspired by human motor behavior, coordinate the fingers.

The first mechanism coordinates the closure of the fingers by distributing the spring stiffnesses according to a geometric scaling law:

$$k_{ij} = k_0 \alpha^{\frac{i-1}{N_f-1}} \beta^{\frac{j-1}{N_p-1}}, \quad (5)$$

where N_f and N_p are the numbers of fingers and phalanges, i is the finger index, j is the phalanx index, and k_0 is a base stiffness chosen experimentally. This law reduces the stiffness design space to two parameters, α and β , which can be explored systematically. The parameter α sets which fingers close first, as illustrated in Fig. 3. The parameter β distributes the attraction among the phalanges of each finger, which sets the closure pattern within a finger.

The second mechanism makes the fingers reach the template surface nearly simultaneously, which prevents ejection at first contact. Distance-dependent damping slows each finger down as it approaches the surface. Each spring is

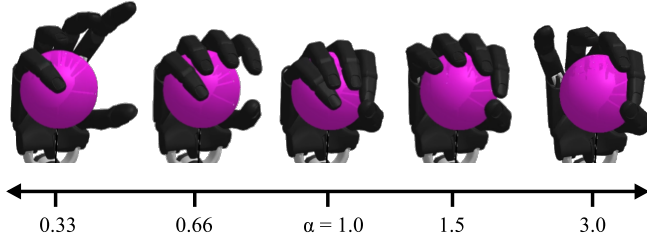


Fig. 3. A single parameter governs the coordination between fingers during closure. At $\alpha = 1$, all fingers close symmetrically. Increasing α advances the thumb and index first, creating a stabilizing pinch before full-hand enclosure; decreasing α advances the outer fingers first.

TABLE I
PARAMETER VALUES FOR EACH GRASP TYPE.

	k_{obj}	c_{max}	k_0	α	β	c_0	c_1	d
Medium wrap	5.0	5.0	0.05	1.5	0.5	0.05	0.1	0.01
Power sphere	5.0	5.0	0.05	1.0	0.1	0.15	0.2	0.01
Lateral pinch	5.0	5.0	0.1	1.0	1.0	0.05	0.1	0.005

paired with a damper whose coefficient follows

$$c(|z|) = c_0 + c_1 \exp\left(-\frac{|z|}{\tau}\right). \quad (6)$$

A uniform term c_0 suppresses oscillations, while an exponential term c_1 decelerates the finger near the template surface. The decay constant is $\tau = -d/\ln(0.2)$, where d is the distance at which the exponential damping reaches 20 % of its maximum value.

The parameters of each spring and damper are adapted to the specific template and align with the human grasp taxonomy. That is, in addition to the sliding-contact architecture, they encode the features of each grasp type (Fig. 2). For the medium wrap, $\alpha > 1$ creates a stabilizing thumb–index pinch before the whole hand closes. For the power sphere, angular springs maintain the spacing between the fingers and promote enclosure, and $\alpha = 1$ makes the fingers close symmetrically, which prevents lateral ejection (Fig. 3). Finally, for the lateral pinch, both extremities of the index finger are driven toward the box corners, maximizing span. Table I lists the parameter values for each grasp type.

IV. EXPERIMENTS

A. Experimental setup

Experiments were conducted with a Shadow Dexterous Hand [10] mounted on a UR10e arm (Fig. 4). The hand is position controlled: its joint controllers track the closing motion computed by the planner. All computations¹ use `VMRobotControl.jl` [49], a Julia library for the design, simulation, control, and optimization of VMC architectures. The test set consists of 27 3D-printed cylinders, spheres, and flat boxes covering the typical sizes of everyday objects.

¹The code, including the grasping pipeline of Section IV, will be released on a GitHub page after the double-blind review process.

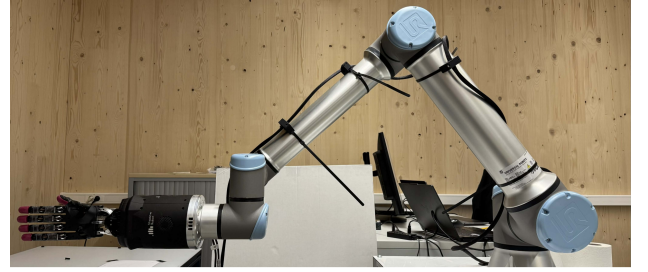


Fig. 4. Robustness tests use 27 3D-printed cylinders, spheres, and boxes (bottom), grasped by a Shadow Dexterous Hand on a UR10e arm (top).

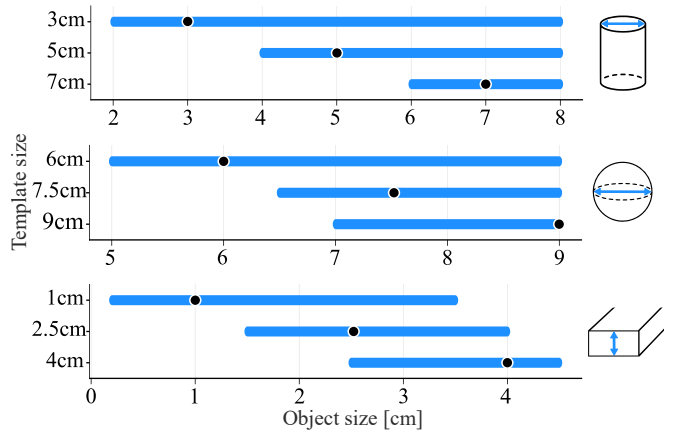


Fig. 5. Size errors of about 1 cm do not compromise grasp success. Blue intervals show the range of successful grasps for each template size; sizes are diameters for cylinders and spheres and thicknesses for flat boxes; black points mark exact matches between template and object.

Cylinders were placed upright, spheres on dedicated stands, and flat boxes on elevated surfaces. For each grasp type, the hand started from a fixed pre-grasp pose relative to the template. A grasp is considered successful if the object is lifted for five seconds without slipping.

B. Robustness to size and pose errors, and comparison

Robustness to size errors was evaluated by choosing three template sizes for each grasp type. The associated grasp type was applied across a range of object sizes typical of human grasping; each grasp was repeated five times, with identical outcomes. Fig. 5 shows the range of object sizes grasped with each template: errors of about 1 cm do not induce any loss of performance. Underestimating the size is tolerated best: the 3 cm cylinder template grasps cylinders up to 8 cm, the largest size tested. Although successful, these mismatched

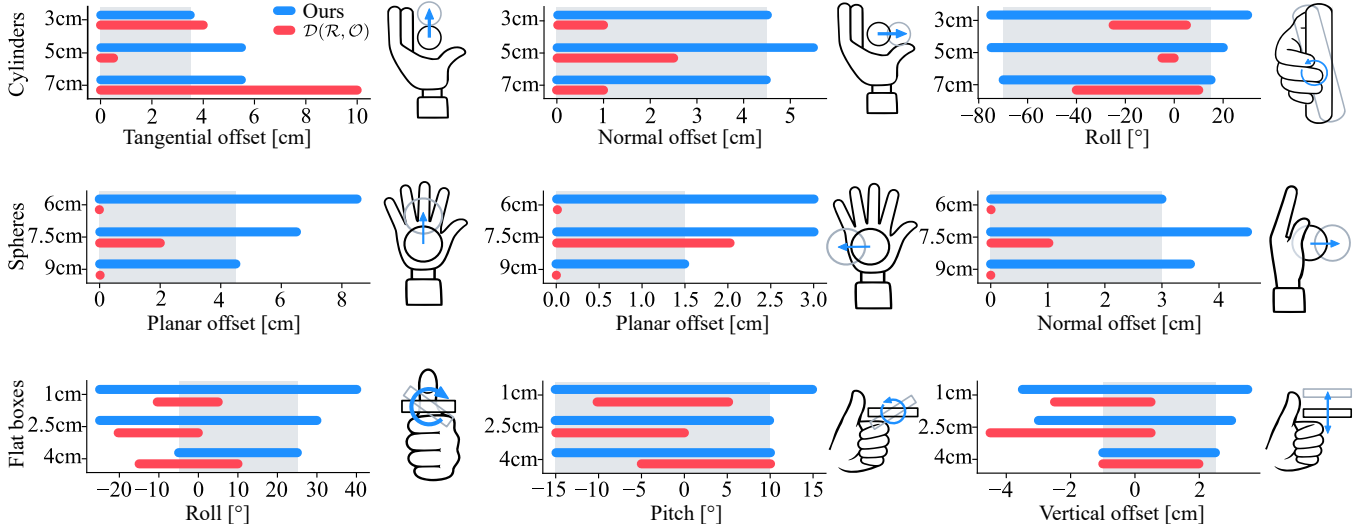


Fig. 6. Our planner (blue) tolerates wider pose errors than $\mathcal{D}(\mathcal{R}, \mathcal{O})$ Grasp (red) in 25 of the 27 axis–object combinations. Bars show the perturbations with successful grasps along three axes per grasp type (icons), for three object sizes; for $\mathcal{D}(\mathcal{R}, \mathcal{O})$, a perturbation counts as successful if at least two of three generated grasps succeed. Grey shaded area marks the range shared by the three objects for our planner.

grasps are less stable and less geometry-adapted than those obtained using a matching template.

Robustness to pose errors was evaluated for each grasp type along three perturbation axes, shown by the hand icons in Fig. 6. Increasing perturbations are applied to the three objects that match the template sizes, with results reported as blue lines. The ranges shared by the three objects (grey shaded ranges) span 1.5 to 4.5 cm in translation and 25 to 85° in orientation.

The red lines in Fig. 6 compare our planner with $\mathcal{D}(\mathcal{R}, \mathcal{O})$ Grasp [11], a state-of-the-art generative grasp synthesis method that infers grasp poses from arbitrary point clouds. $\mathcal{D}(\mathcal{R}, \mathcal{O})$ received ideal point clouds sampled from the object meshes, so both methods start from exact object information before perturbation. Using the same perturbation protocol, three grasp poses were generated per object, and a perturbation is considered successful if at least two grasps succeed. The hand was first moved to the nominal grasp pose and then displaced to the perturbed pose. In the case of $\mathcal{D}(\mathcal{R}, \mathcal{O})$, contact during this motion may drag the object along, which reduces the actual perturbation and can lead to overestimates in the robustness of $\mathcal{D}(\mathcal{R}, \mathcal{O})$. In contrast, our approach starts from an open palm, so no contact occurs. Despite this bias, $\mathcal{D}(\mathcal{R}, \mathcal{O})$ tolerates a narrower range than our planner in 25 of the 27 axis–object combinations; in the two exceptions, tangential offsets of the 3 and 7 cm cylinders, the drag of the object affects the estimate. We also tried to extend the comparison to DexGrasp Anything [50]. However, out of the box, the produced grasps were poor, which prevented a proper comparison.

The gap in performance follows from the design of the two planners: $\mathcal{D}(\mathcal{R}, \mathcal{O})$ generates static grasp poses anchored to specific contact points, which enables complex, object-specific grasps but reduces tolerance to pose errors. In contrast, our planner prescribes no contact points, generates

the whole closing motion, and takes advantage of well-established human grasps. The two approaches are complementary: $\mathcal{D}(\mathcal{R}, \mathcal{O})$ adapts grasps to complex shapes, whereas our planner favors robustness through its reliance on generic templates.

C. Sensitivity to stiffness and damping parameters

We evaluated the sensitivity of the planner to the parameters of the virtual interface on the same nine objects, each matching its template in size and pose. Stiffness and damping were independently varied by up to $\pm 25\%$ from their nominal values (Table I). No grasp failures occurred under damping variations, while stiffness variations caused a single failure at the boundary of the tested range, for the largest sphere. This failure resulted from changing α . For the power sphere, the nominal value of 1.0 makes the fingers close symmetrically while a deviation disrupts the enclosure and ejects the object early, as shown in Fig. 3. Our planner therefore does not require precise tuning within $\pm 25\%$: the only failure came from changing the relative stiffness across fingers, not from the absolute parameter values.

D. Everyday objects

We assessed the ability of our planner to grasp everyday objects. Each object was assigned to a coarse template that matches its dominant dimensions: a sphere for approximately isotropic objects, a cylinder for one dominant dimension, and a flat box for two.

We evaluated this generalization on 35 objects from the YCB dataset [51], primarily from the *Food* and *Kitchen* categories, complemented by 45 additional everyday objects (Fig. 7). Our planner grasped 77% of the YCB objects and 82.5% of all 80 objects, which shows that three coarse templates approximate many everyday objects.

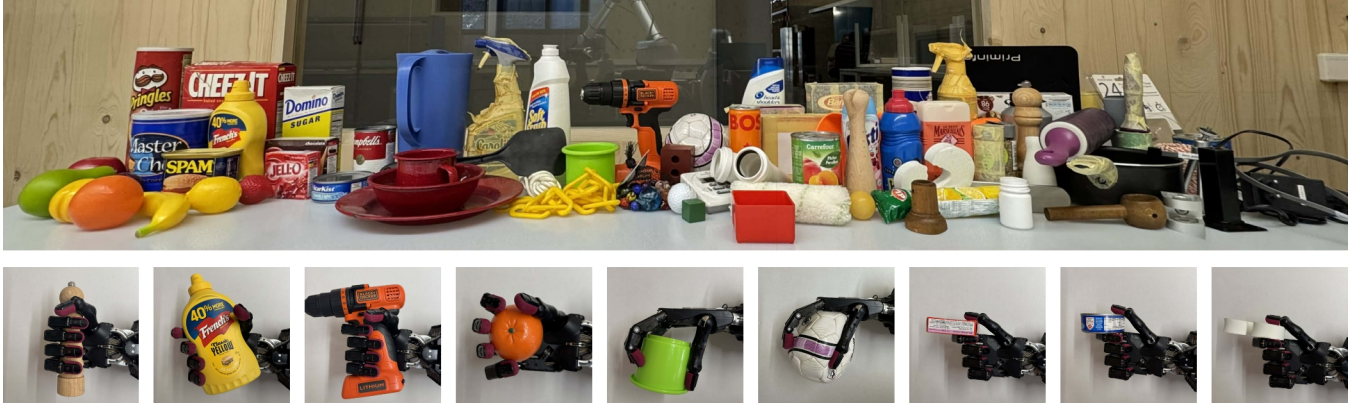


Fig. 7. Three templates suffice to grasp 82.5 % of 80 everyday objects (77 % of the YCB subset). YCB objects appear on the left, additional objects on the right; the bottom row shows example grasps.

As a final validation, we integrated our planner into an autonomous grasping pipeline (Fig. 8, top). An Intel RealSense D435i 3D camera provides a point cloud, from which standard lightweight methods segment the object, fit the template, predict the pre-grasp pose, and reach it; their implementation is included in the released code. Twelve of the 80 objects were tested in five random orientations each, for a total of 60 trials. The planner succeeded in 54 of 60 trials (90 %), which shows that it can be integrated into a conventional grasping pipeline with camera-based sensing.

V. CONCLUSION

This paper shows that the robustness of robot grasping can be improved through proper design of the grasping motion. We have introduced a new motion planner rooted in four key principles and realized via VMC. Without relying on any real-time feedback or adaptation, the planned motions tolerated object size errors of about 1 cm and pose errors of several centimeters and tens of degrees, wider than a state-of-the-art data-driven planner in most cases. Three coarse templates were also enough to grasp 82.5 % of 80 everyday objects, so this robustness does not come at the cost of generalization.

This planner is a first step. It provides desired trajectories which are tracked by the robot hand internal controller. Faster and more dynamic grasps require a proper motion control design, which we will pursue through adaptive feedback and tactile sensing. In a similar fashion, object-template matching is based on simple segmentation and primitive fitting but could be significantly extended through learning mechanisms, allowing for a richer template library, including a library whose templates could be learned through visuotactile sensing.

REFERENCES

- [1] A. Billard and D. Kragic, “Trends and challenges in robot manipulation,” *Science*, vol. 364, no. 6446, p. eaat8414, 2019.
- [2] V. N. Nguyen, S. Tyree, A. Guo, M. Fourmy, A. Gouda, T. Lee, S. Moon, H. Son, L. Ranftl, J. Tremblay *et al.*, “BOP challenge 2024 on model-based and model-free 6D object pose estimation,” *arXiv preprint arXiv:2504.02812*, 2025.



Fig. 8. Within an autonomous pipeline, the planner grasps 12 everyday objects in 54 of 60 trials (90 %). Top: pipeline modules, with our planner outlined in red. Bottom: successes over five trials per object, grouped by grasp type.

- [3] H. Dang and P. K. Allen, “Stable grasping under pose uncertainty using tactile feedback,” *Autonomous Robots*, vol. 36, no. 4, pp. 309–330, 2014.
- [4] T. Feix, J. Romero, H.-B. Schmeider, A. M. Dollar, and D. Kragic, “The grasp taxonomy of human grasp types,” *IEEE Trans. Human-Mach. Syst.*, vol. 46, no. 1, pp. 66–77, 2015.
- [5] I. M. Bullock, T. Feix, and A. M. Dollar, “Finding small, versatile sets of human grasps to span common objects,” in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2013, pp. 1068–1075.
- [6] J. E. Pratt, “Virtual model control of a biped walking robot,” Master’s

thesis, Massachusetts Institute of Technology, Aug. 1995.

- [7] J. Pratt, C.-M. Chew, A. Torres, P. Dilworth, and G. Pratt, "Virtual model control: An intuitive approach for bipedal locomotion," *Int. J. Robot. Res.*, vol. 20, no. 2, pp. 129–143, 2001.
- [8] D. Larby and F. Forni, "Optimal virtual model control for robotics: Design and tuning of passivity-based controllers," *IEEE Trans. Robot.*, vol. 42, pp. 439–454, 2026.
- [9] Y. Zhang, D. Larby, F. Iida, and F. Forni, "Virtual model control for compliant reaching under uncertainties," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2024, pp. 795–801.
- [10] A. Kochan, "Shadow delivers first hand," *Ind. Robot.*, vol. 32, no. 1, pp. 15–16, 2005.
- [11] Z. Wei, Z. Xu, J. Guo, Y. Hou, C. Gao, Z. Cai, J. Luo, and L. Shao, " $\mathcal{D}(\mathcal{R}, \mathcal{O})$ grasp: A unified representation of robot and object interaction for cross-embodiment dexterous grasping," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2025, pp. 4982–4988.
- [12] C. Ferrari, J. Canny *et al.*, "Planning optimal grasps," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, vol. 3. IEEE, 1992, pp. 2290–2295.
- [13] A. Bicchi and V. Kumar, "Robotic grasping and contact: A review," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, vol. 1, 2000, pp. 348–353.
- [14] J. Bohg, A. Morales, T. Asfour, and D. Kragic, "Data-driven grasp synthesis—A survey," *IEEE Trans. Robot.*, vol. 30, no. 2, pp. 289–309, 2013.
- [15] T. G. W. Lum, A. H. Li, P. Culbertson, K. Srinivasan, A. D. Ames, M. Schwager, and J. Bohg, "Get a grip: Multi-finger grasp evaluation at scale enables robust sim-to-real transfer," *arXiv preprint arXiv:2410.23701*, 2024.
- [16] J. Varley, C. DeChant, A. Richardson, J. Ruales, and P. Allen, "Shape completion enabled robotic grasping," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2017, pp. 2442–2447.
- [17] S. Ekvall and D. Kragic, "Learning and evaluation of the approach vector for automatic grasp generation and planning," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2007, pp. 4715–4720.
- [18] D. D. Hoffman and W. A. Richards, "Parts of recognition," *Cognition*, vol. 18, no. 1–3, pp. 65–96, 1984.
- [19] I. Biederman, "Recognition-by-components: A theory of human image understanding," *Psychol. Rev.*, vol. 94, no. 2, pp. 115–147, 1987.
- [20] A. T. Miller, S. Knoop, H. I. Christensen, and P. K. Allen, "Automatic grasp planning using shape primitives," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, vol. 2, 2003, pp. 1824–1829.
- [21] K. Huebner, S. Ruthotto, and D. Kragic, "Minimum volume bounding box decomposition for shape approximation in robot grasping," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2008, pp. 1628–1633.
- [22] K. Huebner and D. Kragic, "Selection of robot pre-grasps using box-based shape approximation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2008, pp. 1765–1770.
- [23] A. Palleschi, F. Angelini, C. Gabellieri, D. W. Park, L. Pallottino, A. Bicchi, and M. Garabini, "Grasp it like a pro 2.0: A data-driven approach exploiting basic shape decomposition and human data for grasping unknown objects," *IEEE Trans. Robot.*, vol. 39, no. 5, pp. 4016–4036, 2023.
- [24] C. Goldfeder, P. K. Allen, C. Lackner, and R. Pelossof, "Grasp planning via decomposition trees," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2007, pp. 4679–4684.
- [25] G. Vezzani, U. Pattacini, and L. Natale, "A grasping approach based on superquadric models," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2017, pp. 1579–1586.
- [26] M. Przybylski, T. Asfour, and R. Dillmann, "Unions of balls for shape approximation in robot grasping," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2010, pp. 1592–1599.
- [27] —, "Planning grasps for robotic hands using a novel object representation based on the medial axis transform," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2011, pp. 1781–1788.
- [28] N. Vahrenkamp, E. Koch, M. Waechter, and T. Asfour, "Planning high-quality grasps using mean curvature object skeletons," *IEEE Robot. Autom. Lett.*, vol. 3, no. 2, pp. 911–918, 2018.
- [29] Q. Lu and T. Hermans, "Modeling grasp type improves learning-based grasp planning," *IEEE Robot. Autom. Lett.*, vol. 4, no. 2, pp. 784–791, 2019.
- [30] D. Lyons, "A simple set of grasps for a dextrous hand," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, vol. 2, 1985, pp. 588–593.
- [31] S. A. Stansfield, "Robotic grasping of unknown objects: A knowledge-based approach," *Int. J. Robot. Res.*, vol. 10, no. 4, pp. 314–326, 1991.
- [32] J. Romero, H. Kjellstrom, and D. Kragic, "Modeling and evaluation of human-to-robot mapping of grasps," in *Proc. Int. Conf. Adv. Robot. (ICAR)*, 2009, pp. 1–6.
- [33] S. B. Kang and K. Ikeuchi, "Toward automatic robot instruction from perception-mapping human grasps to manipulator grasps," *IEEE Trans. Robot. Autom.*, vol. 13, no. 1, pp. 81–95, 1997.
- [34] A. Morales, T. Asfour, P. Azad, S. Knoop, and R. Dillmann, "Integrated grasp planning and visual object localization for a humanoid robot with five-fingered hands," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2006, pp. 5663–5668.
- [35] K. Harada, K. Kaneko, and F. Kanehiro, "Fast grasp planning for hand/arm systems based on convex model," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2008, pp. 1162–1168.
- [36] Z. Deng, G. Gao, S. Frintrop, F. Sun, C. Zhang, and J. Zhang, "Attention based visual analysis for fast grasp planning with a multi-fingered robotic hand," *Front. Neurobot.*, vol. 13, p. 60, 2019.
- [37] Z. Deng, B. Fang, B. He, and J. Zhang, "An adaptive planning framework for dexterous robotic grasping with grasp type detection," *Robot. Autom. Syst.*, vol. 140, p. 103727, 2021.
- [38] H. B. Amor, O. Kroemer, U. Hillenbrand, G. Neumann, and J. Peters, "Generalization of human grasping for multi-fingered robot hands," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2012, pp. 2043–2050.
- [39] J. Lundell, E. Corona, T. N. Le, F. Verdoja, P. Weinzaepfel, G. Rogez, F. Moreno-Noguer, and V. Kyriki, "Multi-finger: Generative coarse-to-fine sampling of multi-finger grasps," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2021, pp. 4495–4501.
- [40] D. Dimou, J. Santos-Victor, and P. Moreno, "Grasp pose sampling for precision grasp types with multi-fingered robotic hands," in *Proc. IEEE-RAS Int. Conf. Humanoid Robots (Humanoids)*, 2022, pp. 773–779.
- [41] Y. Zhang, J. Hang, T. Zhu, X. Lin, R. Wu, W. Peng, D. Tian, and Y. Sun, "Functionalgrasp: Learning functional grasp for robots via semantic hand-object representation," *IEEE Robot. Autom. Lett.*, vol. 8, no. 5, pp. 3094–3101, 2023.
- [42] M. Santello, M. Flanders, and J. F. Soechting, "Postural hand synergies for tool use," *J. Neurosci.*, vol. 18, no. 23, pp. 10 105–10 115, 1998.
- [43] M. Santello, M. Bianchi, M. Gabicini, E. Ricciardi, G. Salvietti, D. Prattichizzo, M. Ernst, A. Moscatelli, H. Jörntell, A. M. Kappers *et al.*, "Hand synergies: Integration of robotics and neuroscience for understanding the control of biological and artificial hands," *Phys. Life Rev.*, vol. 17, pp. 1–23, 2016.
- [44] M. Ciocarlie, C. Goldfeder, and P. Allen, "Dexterous grasping via eigengrasps: A low-dimensional approach to a high-complexity problem," in *Proc. Robot.: Sci. Syst. (RSS), Workshop on Sensing and Adapting to the Real World*, 2007.
- [45] R. Ortega, A. J. van der Schaft, I. Mareels, and B. Maschke, "Putting energy back in control," *IEEE Control Syst. Mag.*, vol. 21, no. 2, pp. 18–33, 2001.
- [46] M. W. Spong, "An historical perspective on the control of robotic manipulators," *Annu. Rev. Control Robot. Autom. Syst.*, vol. 5, pp. 1–31, 2022.
- [47] N. Chopra, M. Fujita, R. Ortega, and M. W. Spong, "Passivity-based control of robots: Theory and examples from the literature," *IEEE Control Syst. Mag.*, vol. 42, no. 2, pp. 63–73, Apr. 2022.
- [48] C. Secchi, S. Stramigioli, and C. Fantuzzi, *Control of Interactive Robotic Interfaces: A Port-Hamiltonian Approach*, ser. Springer Tracts in Advanced Robotics. Berlin: Springer, 2007, no. volume 29.
- [49] D. Larby, "VMRobotControl.jl," <https://github.com/Cambridge-Control-Lab/VMRobotControl.jl>, 2024, accessed: Mar. 2026.
- [50] Y. Zhong, Q. Jiang, J. Yu, and Y. Ma, "Dexgrasp anything: Towards universal robotic dexterous grasping with physics awareness," in *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2025, pp. 22 584–22 594.
- [51] B. Calli, A. Singh, A. Walsman, S. Srinivasa, P. Abbeel, and A. M. Dollar, "The YCB object and model set: Towards common benchmarks for manipulation research," in *Proc. Int. Conf. Adv. Robot. (ICAR)*, 2015, pp. 510–517.