Grasping and Object Reorientation for Autonomous Construction of Stone Structures

Martin Wermelinger, Ryan Johns, Fabio Gramazio, Matthias Kohler, and Marco Hutter

Abstract—**Building large and stable structures from highly irregular stones is among the most challenging construction tasks with excavators. In this letter, we present a method for grasp planning and object manipulation that enables the world's first autonomous assembly of a large-scale stone wall with an unmanned hydraulic excavator system. Our method utilizes point clouds and mesh surface reconstruction of stones in order to plan grasp configurations with a 2-jaw gripper mounted on the excavator. Besides considering the geometry of the stone to sample force closure grasps, the grasp planner also takes into account collision constraints during object pick and place, informed by a LiDAR based point cloud map. Furthermore, we show an approach to reorient arbitrarily shaped objects that are not feasible to be directly placed at the desired location without violating collision constraints. Using a physics engine, we find a settled intermediate pose that allows direct placement and is reachable from the initial stone pose. The applicability of the proposed grasp planning method is demonstrated with the construction of a dry stone wall composed of over one hundred boulders using an autonomous excavator. We show a high primary grasp success rate (82.2 %) and illustrate how the system recovers from slippage by relocating the object and re-planning the grasp correspondingly.**

Index Terms—**Robotics and automation in construction, grasping, manipulation planning.**

I. Introduction

Automating construction has gained increased popularity in robotics recently as it promises more efficient, more sustainable, and safer building operations [1] and enables the implementation of novel types of architectural structures with unprecedented complexity and functional properties [2]. In our work we investigate the direct utilization of raw or minimally processed material that is either naturally present on the construction site or can be easily provided in large amounts such as natural stones or rubble. Locally sourced material and autonomous construction machines have a great potential for landscape construction in dangerous-to-access and remote places (like mountainous terrain or disaster sites), while the incorporation of natural or waste material into the robotic construction process reduces their ecological footprint [3]. This is particularly relevant for the task of building utility structures (e.g. noise protection walls, river bank reinforcements, coastal protections or retaining walls) that usually consist of great quantities of irregular elements, and where the global approximate shape is more important than the exact composition. Current on-site robotic systems focus on using regular building material [4], whereas we focus on construction with irregularly shaped rigid objects. We showed in our recent work [5] how sampling-based methods can be used to develop assembly plans for constructing large-scale dry stone walls, like retaining walls, from unprocessed stones. The aforementioned wall planner determines stable poses for available stones to achieve a specified wall shape, combining heuristics from historic dry stone masonry with methods for shape matching, iterative alignment, and physics-based settling. To execute this plan, the biggest challenge is a reliable manipulation of such irregularly shaped objects including the grasping and reorientation necessary to place them into tightly constrained locations within an assembly. In this letter we present an approach for sampling grasp configurations of an excavator-mounted 2-jaw gripper based on a point cloud combining reconstructed object meshes and a LiDAR based map. We evaluate the force closure of grasps using a learning-based classifier, while also considering collision constraints at both ends of the pick and place sequence, in order to assess the feasibility of relocating an object to the desired location. Furthermore, we include a

Fig. 1. The autonomous excavator HEAP [6] assembling a dry stone retaining wall with irregularly shaped boulders. As shown in the inset, an arm mounted LiDAR is used for mapping and stone localization.

Manuscript received December 4, 2020; accepted March 1, 2021. Date of publication March 31, 2021; date of current version April 21, 2021. This letter was recommended for publication by Associate Editor Gianluca Garofalo and Editor Youngjin Choi upon evaluation of the reviewers' comments. This work was supported in part by the Swiss National Science Foundation through the National Centre of Competence in Digital Fabrication (NCCR dfab). *(Corresponding author: Martin Wermelinger.)*

Martin Wermelinger and Marco Hutter are with Robotic Systems Lab, ETH Zurich, 8050 Zürich, Switzerland (e-mail: johns@arch.ethz.ch; johns@arch.ethz.ch).

Ryan Johns, Fabio Gramazio, and Matthias Kohler are with Gramazio Kohler Research, ETH Zurich, 8005 Zurich, Switzerland (e-mail: martiwer@ethz.ch; gramazio@arch.ethz.ch; kohler@arch.ethz.ch).

This letter has supplementary downloadable material available at https://doi.org/10.1109/LRA.2021.3070300, provided by the authors.

Digital Object Identifier 10.1109/LRA.2021.3070300

method for reorienting objects if no direct placement is feasible that finds a stable intermediate placement pose by letting the mesh reconstruction of the object settle in a physics simulation and verifing that collision-free grasp configurations are available for placement. Handling large-scale objects poses several challenges that need to be addressed. Conventional approaches for assembling complex-shaped objects in a robotics cell focus on clearly defined geometries [7], simplifying the grasp planning and the elimination of alignment errors, whereas we handle arbitrary shaped objects without simple geometric features like straight edges. In addition, the use of an automated excavator limits the end-effector range of motion compared to an articulated robot arm, which needs to be incorporated in the grasp planning and object reorientation method. The proposed approach enabled construction of the world's first large-scale wall using an autonomous excavator, whereby we manipulated more than a hundred boulders that each weigh several hundred kilograms and have a unique and highly diverse geometry (see Fig. 1). We show a detailed evaluation of the grasp and reorientation success rate and discuss slippage cases and possible improvements. To the best of our knowledge, this is the first demonstration of object grasp planning and reorientation with irregularly shaped objects at this size for assemblies on an architectural scale.

II. Related Work

Grasp planning is often divided into *analytical* and *empirical* (data-driven) methods. Analytical methods evaluate the performance of a grasp according to physical properties such as stability, equilibrium, dexterity, and dynamic behaviour but only under the assumption of having the exact object model and its pose in the scene [8]. Grasps can be synthesized from a constraint optimization problem over one or several measures of the mentioned properties that is hard to solve in case of our non-convex objects. Empirical methods on the other hand rely on sampling grasp configurations and ranking them according to metrics that typically come from simulation trials, physical trials or human labels [9]. They place more weight on the perceptual object representation and better accommodate uncertainties in perception and execution, such as those present in our case of reconstructing and localizing irregularly shaped objects. There exist multiple empirical methods that rely on grasp detection approaches to generate grasp configurations in cluttered scenes directly, without the need of first localizing single object instances [10]–[12]. These approaches begin by generating a large number of grasp candidates on the input scene point cloud, usually originating from an RGB-D camera. They then evaluate the probability of the candidates being a grasp, e.g. in terms of force closure using a classifier or regression system trained on a large amount of labeled data. These methods generalize well for new objects, as they detect grasps based on graspable regions independent of object instances, but are prone to find contact points that are spread across multiple objects. In our outdoor case, depth images of the scene are not directly available. Therefore, we perform the grasp detection on the point cloud generated by 3D LiDAR mapping. We combine the LiDAR based map with reconstructed object meshes to complement the point cloud used for grasping, as well as to ensure that contact points are only located on the object of interest. To sample grasp hypotheses on the point cloud and classify the grasp candidates, we use a similar approach as presented in [13]. However, from the sampled grasps we perform a further selection step where the remaining candidates are filtered for heuristic criteria derived from the specific task to obtain the final grasp candidate. The idea of filtering the best generated hypotheses is that we are not necessarily interested in the one optimal grasp [14], but we are rather interested in a sufficiently good grasp that fulfills the task specific criteria.

Object reorientation is the task of moving an object between different 3D orientations to make it accessible for further manipulation tasks such as assembly. Traditional methods use pick-and-place motions, where the manipulator grasps the object firmly and rotates it to the desired stable pose [15]. This process may repeat several times, depending on the robot's range of motion and workspace limitations, making it necessary to plan the motion sequencing resulting in a constraint satisfaction problem [16]. To avoid time consuming motion sequence planning, we design a reorientation sequence that allows the object to be brought to the desired orientation for placement within in a single reorientation motion, taking advantage of the gripper's ability to rotate continuously. An alternative method to use less gripper and arm motion in constrained workspaces is pivoting [17], an object is pinch-grasped and can passively rotate about the grasp axis while it remains in contact with the ground surface during motion, allowing the object to reorient around an arbitrary edge and to decouple object rotation and gripper motion. Pivoting works best for regular, prismatic objects which is not the case in our application, and would require multiple pivoting motions. Furthermore, the gripper has to be able to switch between a pivoting grasp and a firm grasp, e.g. by changing gripping force or the finger-object contact geometry [18], which is not possible for the hydraulic gripper of the excavator. More dynamic motion primitives for object reorientation such as throwing and catching [19] or in-hand pivoting with inertial forces [20], [21] have only been presented for simple object geometries with clearly defined rotation axes and are not suitable for rotating objects with the target size and weight.

III. Grasp Pose Planning

The goal of the grasp pose planning is to find viable grasp configurations in order to pick an object and place it at the desired location in the assembly structure. In the case of the autonomous excavator, a grasp configuration is defined as a 6 DoF pose of the gripper where a contact configuration with the object can be performed. We aim to find *force closure* grasps, which are grasp configurations with contact wrenches that span the centroid of the object. We will focus here on finding grasp poses for 2-finger grippers, as the excavator is equipped with a 2-jaw angular gripper, consisting of mechanically coupled jaws and a palm. For collision checking between the gripper and planning point cloud, the gripper shape is approximated with convex polyhedra that encompass the jaws and palm. This collision check requires the object mesh to be completely reconstructed beforehand. For reconstruction we segment the point cloud of the excavator's surrounding to identify object instances, perform an initial grasp, and scan the object using excavator mounted LiDARs [22]. The grasp pose planning consists of three steps: generation of a grasp

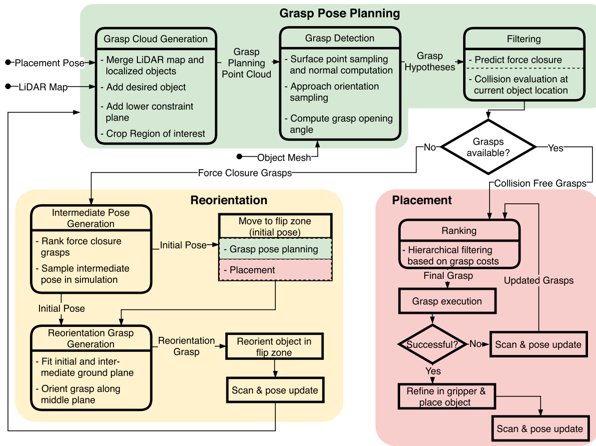

Fig. 2. The grasp pose planning and object reorientation pipeline.

point cloud; detection of grasp candidates on the grasp point cloud; and subsequent filtering and ranking of the candidates to obtain a feasible grasping configuration. If none of the sampled grasp configurations are valid at the pick location, a reorientation of the object (Section IV) is attempted. An overview of the grasp pose planning and reorientation procedure is depicted in Fig. 2.

A. Grasp Planning Point Cloud Generation

The collision detection of possible grasp configurations is done directly on a point cloud representation of the scene, called the *grasp planning point cloud*. Grasp configurations are sampled at the desired placement location in the structural assembly, as ultimately the grasp has to be designed such that the placement is feasible.

1) LiDAR Mapping: To obtain the grasp point cloud, we create a map of the excavator's surrounding using a LiDAR sensor mounted on the stick (second link of the excavator arm, see Fig. 1). LaserSLAM [23], a laser-based graph SLAM is used to create a consistent map over long time- and motion intervals. It considers the odometry constraints provided by the excavator's state estimation, fusing Global Navigation Satellite System (GNSS) measurements with the Inertial Measurement Unit (IMU), and scan matching constraints provided by registering a new scan pair using the Iterative Closest Point (ICP) algorithm with a map consisting of previous scans. To prevent map corruption by self-see points, the excavator's arm or legs in the LiDAR scans are filtered to get a 3D point cloud that only includes the local surrounding of the robot. By using an arm mounted LiDAR sensor, it is possible to physically adjust its view point to see the back of objects as well as scan regions in the map which are higher than the robot location itself, which is necessary to build tall structures like walls.

2) Object Models and Ground Plane: Even with a LiDAR that can change its view point, the scene can still be incomplete due to occlusion, especially within the assembly itself. Therefore we augment the map point cloud with point clouds of the object models that are localized in the scene, and of the desired object at placement. We compute the desired motion direction for object placement, called desired placement direction $\boldsymbol{n}_{\text{place, des}}$ (see Fig. 3(a)), by conducting ray casting from the object centroid

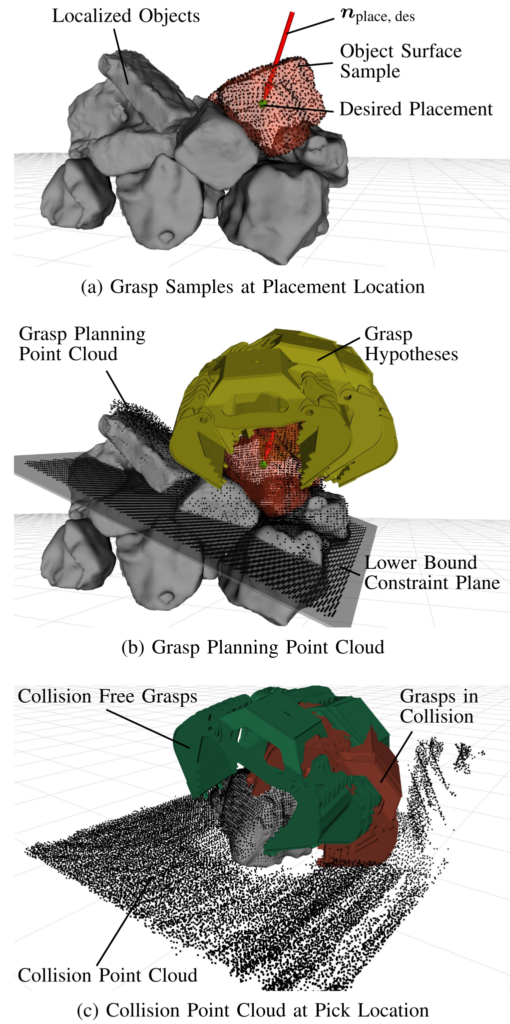

Fig. 3. To place an object at a desired location in the assembly structure (red), the desired placement direction $\boldsymbol{n}_{\text{place, des}}$ with highest margin is computed. Object surface points are sampled to get the surface normal as possible grasp approach directions (a). Grasp hypotheses are generated on the grasp planning point cloud (b) and reprojected to the object pick location for collision detection (c).

at the desired pose (green sphere). Rays that to do not hit the ground or surrounding objects are averaged to obtain the placement direction with highest margin to obstacles. A lower bound constraint plane orthogonal to the desired placement direction is added at the lowest point of the desired object surface in the placement direction. The idea of the constraint plane is that points below this plane can be cropped and ignored, reducing the size of the point cloud significantly and speeding up the collision detection. Finally, the grasp planning point cloud is cropped to a box region of interest around the desired object position with an edge length that is twice the maximum gripper aperture of 1.8 m (see Fig. 3(b)).

B. Grasp Detection

The intention of grasp detection is to generate a large number of grasp hypotheses on an object. A grasp hypothesis is valid if the gripper does not collide with the environment, verified by intersecting the gripper model with the grasp planning point

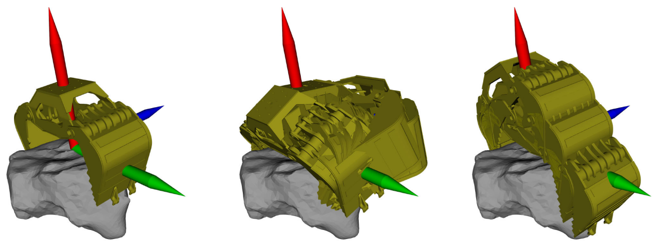

Fig. 4. A local reference frame $\mathcal{F}$ is assigned to a surface sample point with the x-axis pointing away from the object surface (a). Multiple orientations are generated for each sample point by rotating the local frame around the y-axis (b) and z-axis (c) in discrete intervals.

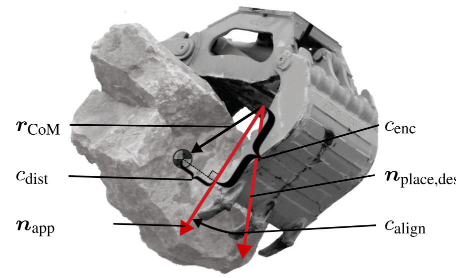

Fig. 5. The grasp cost is dependent on the distance of the the gripper palm to the centroid $\boldsymbol{r}_{\text{CoM}}$ and the alignment of the approach direction $\boldsymbol{n}_{\text{app}}$ to the desired placement direction $\boldsymbol{n}_{\text{place, des}}$.

cloud. To generate the grasp hypotheses, N points are sampled on the point cloud representation of the desired object O_i (see Fig. 3(a)). Each sample p is assigned a local reference frame $\mathcal{F}$ by evaluating the Eigenvectors of the matrix

$$\boldsymbol{M}(p) = \sum_{q \in \boldsymbol{B}_r(p)} \boldsymbol{n}(q)\boldsymbol{n}(q)^\top \tag{1}$$

where $\boldsymbol{n}(q)$ is the outwards pointing unit surface normal at point q, and $\boldsymbol{B}_r(p)$ is the r-ball around point p. The local reference frame is composed of $\mathcal{F} = [\boldsymbol{\nu}_3(p), \boldsymbol{\nu}_2(p), \boldsymbol{\nu}_1(p)]$, where $\boldsymbol{\nu}_1(p)$ corresponds to the largest Eigenvalue and $\boldsymbol{\nu}_3(p)$ to the smallest one. This assures that the x-axis of the reference frame is pointing away from the object surface, and the z-axis is pointing along the axis of minimal curvature.

For each sample point p, we generate multiple grasp orientations by rotating the local reference frame around the y- and z-axis in discrete intervals (see Fig. 4). The inverse x-axis of the rotated grasp orientation is denoted as approach direction $\boldsymbol{n}_{\text{app}}$ of the gripper. A convex polyhedral gripper collision model is moved along the approach direction, with several opening angles, until palm or jaws are in contact with the grasp planning point cloud. We add a sampled gripper configuration to the list of grasp hypotheses if the closing region of the jaws only contains points of the desired object O_i and is not empty (see Fig. 3(b)). With this grasp detection approach we can sample grasps only on the object of interest whilst ensuring that there are no undesired contacts with other objects or surrounding during placement.

C. Grasp Filtering and Ranking

The grasp detection generates hundreds of grasps for a single object that are filtered and ranked according to their applicability and likelihood of success.

1) Force Closure: In a first step, the grasp hypotheses are evaluated for force closure. We use a grasp classifier using a four-layer Convolutional Neural Network (CNN) presented in [13] that predicts if a grasp is a force closure based on the planning point cloud and a 2-finger gripper model. This classifier is applied to reduce the detected grasp hypotheses, predicting whether the grasps are force closure—however many candidates may still be valid.

2) Collision at Pick Location: Subsequently, the valid grasps and the desired placement direction are projected to the actual location of the object O_i in the scene to check whether they are also valid for picking the object. Similar to the grasp pose planning point cloud, the LiDAR map around the object is augmented with the localized objects and used to evaluate collision of the grasp hypotheses at the pick location (see Fig. 3(c)). All colliding grasps are discarded, and if none of the detected grasps are collision free at the pick location, a reorientation of the object (see Section IV) is necessary to achieve the desired placement.

3) Grasp Cost: The remaining collision free grasps are ranked and filtered by task specific criteria that proved to provide reliable grasps. First the grasps are ranked according to their alignment with the approach direction. Given the desired placement direction $\boldsymbol{n}_{\text{place, des}}$ and the approach direction of a grasp hypothesis (the normal pointing away from the palm) $\boldsymbol{n}_{\text{app}}$ the alignment cost c_{align} is given as the angle between the two vectors

$$c_{\text{align}} = \arccos\left(\boldsymbol{n}_{\text{place, des}}^\top \boldsymbol{n}_{\text{app}}\right). \tag{2}$$

We prefer grasp approach directions aligned with the desired placement direction because this gives the most margin with respect to collision. The 50 % grasps with the lowest alignment cost c_{align}, or a minimum of 40, are selected and ranked on how much the object is enclosed by the gripper. An enclosing (or power) grasp is preferable to a pinching (or precision) grasp as it is more likely to withstand disturbances. The enclosing cost c_{enc} is represented by the distance between the palm and the centroid (CoM) in the approach direction

$$c_{\text{enc}} = \boldsymbol{r}_{\text{CoM}}^\top \boldsymbol{n}_{\text{app}}, \tag{3}$$

where $\boldsymbol{r}_{\text{CoM}}$ is the position vector from the palm center to the object centroid $\boldsymbol{c}_i$. Again, the 50 % grasps with the lowest cost, but a minimum of 20, are selected. Finally, we minimize the closeness of the grasp to the centroid of the object and select the best grasp based on the cost c_{dist} for execution. The closeness of a grasp to the centroid is given by the distance

$$c_{\text{dist}} = \frac{\|\boldsymbol{r}_{\text{CoM}} \times \boldsymbol{n}_{\text{app}}\|}{\|\boldsymbol{n}_{\text{app}}\|} \tag{4}$$

between the centroid of a line going through the Tool Center Point (TCP) along the approach direction and the centroid. This criterion is motivated by the need to avoid torsional moment on the grasped object during motion that might lead to rotational shift of the object in the gripper. The measures of the grasp costs are illustrated in Fig. 5.

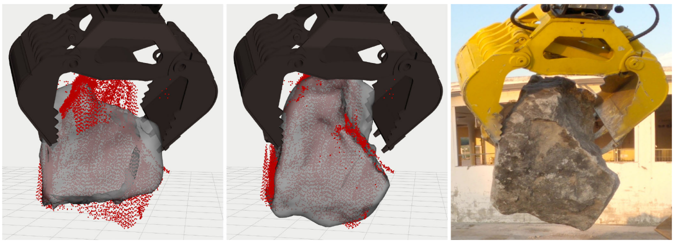

Fig. 6. In-hand pose refinement: The gripper is rotated in front of the cabin mounted LiDARs to generate a scan (red) of the grasped stone (grey). An ICP update step is used to align the in-hand stone pose to the scan.

D. In-Hand Pose Refinement and Failure Recovery

As the gripper is closing symmetrically and not adapting to the contacts, the stone might slightly move with respect to the gripper during the closing and lifting (while still maintaining the grasp). To increase the execution accuracy, the stone pose in the gripper is updated while moving to the placement location. During the cabin swing motion, the gripper is rotated to scan the grasped stone with the cabin roof mounted LiDARs, and the stone is re-registered with respect to the gripper using an ICP update (see Fig. 6).

If the stone slips during the gripper closing and the grasp cannot be maintained, the arm is lowered and the gripper opened again to refine the stone's pose in world frame by running an ICP update. The grasp with the next-best rank is executed to recover from slippage and to proceed with stone placement.

IV. Object Reorientation

In order to allow for maximum flexibility in the online structural assembly planning, the assembly planner has no notion of the prior object orientation, meaning that it may request to place the objects in an arbitrary orientation. As such, it might not be possible to find a grasp pose that allows the object to be moved directly from its current pose to that desired location, due to the excavator's limited range of motion, or more typically because of collision constraints at the pick and place locations. In this case, an intermediate reorientation of the object is necessary to bring the object from its current pose to a configuration that enables it to find a grasp pose suitable for placement.

A. Intermediate Pose Generation

First, we find a nominal desired grasp configuration at the desired placement pose by ranking all force closure grasp according to the grasp cost presented in Section III-C3 (including grasps with collision at pick location) and use the best ranked grasp. We preferably want to reorient the object to an intermediate pose such that the desired grasp approaches the object from the top, guaranteeing to be in the motion range of the excavator and having the smallest chance of occlusion. To find a physically consistent and stable intermediate object pose, the object is put in a physics simulation such that the placement direction of the desired grasp is perpendicular to the ground and the object is slightly above the ground. The physics simulation is stepped forward to let the object settle on the ground and it is verified whether the desired grasp pose is still feasible, i.e. collision free and reachable. If the desired grasp pose is not feasible on the settled object, the initial orientation in the simulation is slightly perturbed to obtain a different settled pose. This is repeated until a feasible settled object pose is achieved that serves as intermediate object pose. The reorientation process preferably takes places on even ground to avoid collision and in our case the simulation uses flat ground as well, however work for arbitrary ground as long as it is mapped and represented in simulation.

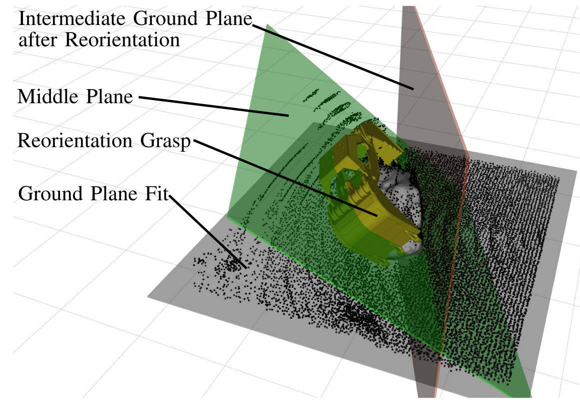

Fig. 7. The reorient. grasp is aligned along the middle plane between the current ground plane and the desired intermediate ground plane after flip.

B. Reorientation Grasp Generation

To move the object from its initial pose to the intermediate pose, a designated grasp is needed that respects the collision constraints from the initial and the intermediate pose. Namely, the surrounding ground planes of the initial and intermediate pose are quite restrictive collision constraints that mutually exclude many possible solutions found by the grasp pose planning pipeline presented in Section III and enforce grasp configurations that are aligned with both ground planes. Therefore, we compute the desired grasp configuration based on the ground planes at the initial and intermediate object pose (see Fig. 7). First, we project the intermediate ground plane with respect to the object frame to the initial object pose. Then, the approach direction of the reorientation grasp is oriented along the middle plane between the initial ground plane and the projected intermediate ground plane, and the grasp is aligned such that the gripper jaws move along the middle plane. Finally, the open polyhedral gripper model is moved along the approach direction until there is contact with the ground planes or the object, and the gripper jaws are closed until they are in contact with the object.

C. Flipping Zone

The reorientation grasp approach direction is inherently located within the angular domain between the ground plane and a maximum tilt of $\frac{\pi}{4}$ rad from the ground plane. Because of the limited motion range of the excavator's joints, such an end-effector configuration can only be reached if the object is relatively close to the chassis. In order to facilitate the reorientation maneuver, the object of interest is therefore first moved to a predefined *flipping zone* on the side of the excavator's

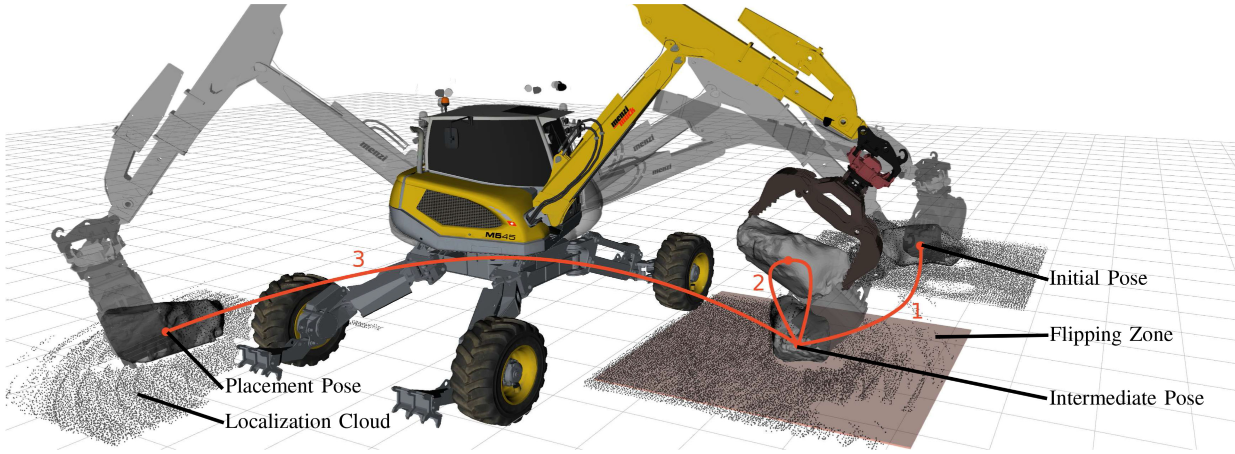

Fig. 8. An object that has to be reoriented in order to be placed, is moved from its initial pose to the *flipping zone* close to the excavator's chassis (1). A regrasp is performed that allows to rotate the object to its desired intermediate orientation (2). Before moving it to its final placement pose, another regrasp is necessary that considers the collision constraints at placement (3). The object is localized each time it is moved using ICP registration of the object mesh to the LiDAR based map.

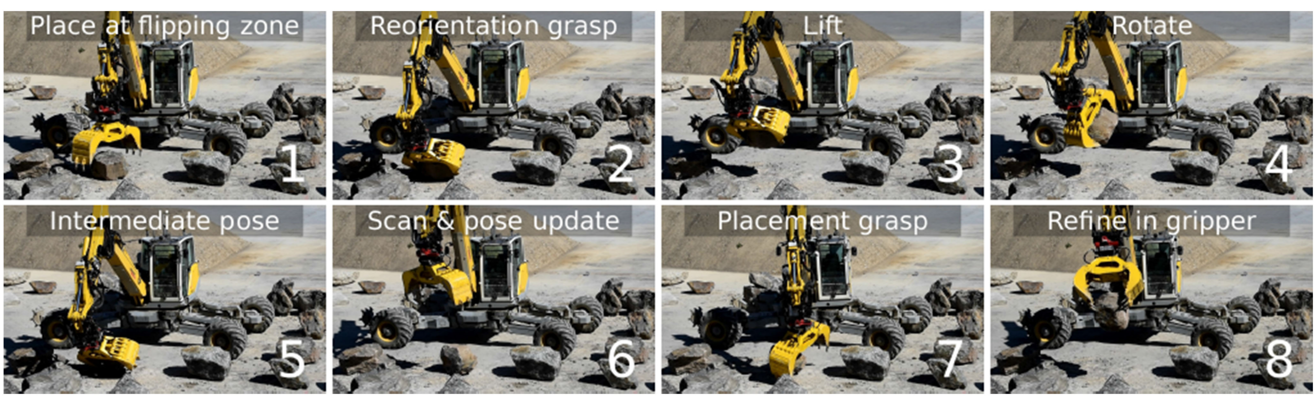

Fig. 9. Object reorientation maneuver of a stone in the flipping zone, from the execution of the reorientation grasp, over the intermediate pose, to the final grasping for placement.

chassis and rotated around the z-axis in world frame, such that the approach direction of the reorientation grasp is pointing towards the chassis (see Fig. 8). After moving the object to the flipping zone, the reorientation grasp is executed and the object is lifted to a height of 1.5 m above ground. It is finally rotated and placed at the same position but with the orientation of the intermediate pose. Each time the object is relocated, its pose in the world frame is refined by running an ICP update using the arm-mounted LiDAR to compensate for deviations that occur during the grasping or settling of an object. With the reoriented object pose, the grasp pose planning pipeline is rerun to obtain a grasp pose for placing the object in the desired assembly location. Fig. 9 shows the reorientation maneuver of an object placed in the flipping zone, from the execution of the reorientation grasp to the final grasping for placement.

V. Experiments

The experiments were conducted using HEAP (Hydraulic Excavator for an Autonomous Purpose), a highly customized Menzi Muck M545 12t walking excavator that was developed for autonomous applications and advanced teleoperation [6]. The fully-mobile 25 DoF excavator is capable of lifting objects to a height of 9 m meters, and can freely manipulate items weighing up to 3000 kg. Cabin- and chassis-mounted IMUs track machine orientation for state estimation, while global cabin localization is achieved by a Leica iCON iXE3 with two GNSS antennas and a receiver obtaining real-time kinematic (RTK) corrections from base stations for improved accuracy. The position and velocity of the hydraulic arm cylinders are measured with draw wire encoders. A LiDAR mounted on the arm (near the gripper) is used for mapping and localization of the objects, as it provides top view of the available material and in-progress structure. Two additional LiDAR scanners are placed at the front edge of the cabin's roof to provide 3D scans for reconstruction and in-hand refinement of the objects.

A. Experiment Description

The applicability of the proposed grasp approach is directly shown in the construction of a large-scale double-faced dry-stone wall with dimensions approx. 10 m $\times$ 4 m $\times$ 2 m (W $\times$ H $\times$ D). To find stone place locations that comply with the target wall geometry and are structurally stable, the geometric

TABLE I
GRASP ATTEMPTS SUCCESS RATE

# Grasps	Succ. Grasps	Slip
443	364 (82.2 %)	79 (17.8 %)

TABLE II
DISTRIBUTION OF SUCC. GRASP ATTEMPTS

# Succ. Grasps	Placements in Wall	To Flip. Zone	Reorient. Grasp	OoW Placement
364	145 (39.8 %)	76 (20.9 %)	77 (21.2 %)	66 (18.1 %)

TABLE III
DISTRIBUTION OF GRASP ATTEMPTS THAT LEAD TO PLACEMENT

Placements in Wall	direct	direct w/o slip	from Reorient.	Reorient. w/o slip
145	69	42 (60.9 %)	76	32 (42.1 %)

TABLE IV
NORMALIZED COSTS (MEAN AND STANDARD DEVIATION) OF PLACEMENT AND SLIP GRASP

Cost	Placement Grasps	Slip
c_{align}	0.9979 ± 0.7646	1.3152 ± 1.0986
c_{enc}	1.0297 ± 0.3182	1.1705 ± 0.2860
c_{dist}	1.1546 ± 0.7966	1.3108 ± 0.7833

assembly planner presented in [5] is used. The wall planner solves for stable stone poses, using the reconstructed surface mesh, in a multistage process involving shape matching, iterative alignment to the target shape, and physics-based settling in combination with heuristics adapted from traditional manual dry stone masonry. The geometric planner provides the desired position and orientation of the stones in the target structure, as well as a placement direction corresponding to the least occlusion. The stones used for construction are irregularly shaped boulders in the size of 445-2425 kg (average 1030 kg). During the several-day construction process, stones were regularly dumped and spread in the excavator's close vicinity, where they could be grasped, scanned and digitally reconstructed for assembly and grasp planning. The task of the grasp pose planning was to find a viable grasp pose that enabled picking a desired stone and placing it at the planned pose without colliding with the existing wall.

In the course of this large-scale experiment, a single wall was built: it required 145 stone placements in the wall, from which 76 object reorientations (52.4 %) were required because no grasp pose could be found that fulfilled the collision constraints at the pick and place locations. The high number of required reorientations is due to the fact that the assembly planner is presently not considering the current orientation of an object, such that it has maximum flexibility in finding viable solutions for the structure.

For every grasp, 2000 surface points were sampled with 10 grasp orientations at each point. The grasp planner took an average of 16.08 ± 2.26 s to run on an Intel Xeon E3-1505 M octa-core 2.8 GHz CPU: 11.62 s for generating the grasp hypotheses, 4.04 s for classifying force closure, and 0.42 s to evaluate collision at pick location and cost computation.

B. Grasp Success Evaluation

To perform the 145 placements in the wall, a total number of 443 grasp attempts were executed with 364 grasps (82.2 %) being labeled as a success (meaning that the stone could be lifted and placed at a new location), whereas 79 slipped out of the gripper during the grasp attempt (see Table I). To place a stone in the desired pose, usually multiple grasps had to be performed, and the 364 successful grasps are distributed as follows: In 145 cases the stone was placed in the wall, 76 grasps were performed to place the stone in the designated flipping zone and 77 grasps were used to reorient the object (see Table II). In 66 cases the stone was grasped successfully, but placed outside of the wall structure (OoW). We indicate this scenario separately, as the grasp was successful (meaning that the stone could be lifted and relocated), but the placement could not be directly executed. This could be due to potential collision at placement because the stone moved during the gripper closing or also if the placement was considered by the human observer to lead to a high risk of partial wall collapse. In this case, the stone was placed again on the ground to replan and adjust the grasp for subsequent placement. In Fig. 10 a selection of the executed placement grasps at the desired stone location is shown.

Table III compares the occurrence of slippage for stones that could be placed directly or had to be reoriented. We can see that since reorientation requires multiple grasps, the chance of slippage increases and the share of placement sequences without slippage is lower (42.1 %) than in the case of direct placement (60.9 %). Through the possibility of updating the object pose in world frame and regrasping when slippage appears, all 145 placement sequences could be carried out in the end with an average of 3.06 grasp attempts. However, the process of updating poses and regrasping is time consuming, and decreases the assembly rate, which was approx. 3 m^2 of wall per day. For comparison, trained dry stone masons achieve to construct 1-2 m^2 of wall per day, but this number highly depends on the available stone size and type [24]. A human operator could increase the construction speed itself, as the robot was operated at reduced speed for safety reasons. But creating an assembly with irregular objects of this size manually is extremely challenging and not practicable as a trial-and-error approach for finding form fits

In Table IV, we report the average hierarchical filter cost terms. The average cost terms are slightly lower for successful placements than in case of slippage, indicating that the cost terms point in the right direction. However, the high standard deviation of the cost terms shows that they are not distinctive enough to ultimately predict grasp success. It is important that a grasp approach, as in our case, is including slippage detection and regrasping to achieve a high placement success rate. The reorientation grasps itself had significantly higher average alignment cost c_{align} (3.0347 ± 0.9603) as they are inherently tilted against the placement direction.

C. Slippage Reasons

Grasping irregular objects is a challenging task, as slight deviation in pose or surface shape lead to varying grasp quality. The contact quality depends highly on local shape features like small dents and protrusions, and these features have a large

Fig. 10. Examples of stone placements at the desired location. The placement grasps comply with the challenging collision constraints and enable to fit objects even in tight spaces.

impact on the grasp success. The accuracy of the contact points is especially influenced by slight asymmetries of the grasp, which can cause the object to move during gripper closing, thus decreasing the success rate. Another common cause of slippage were cone shaped objects that would be pushed out of the gripper by the closing force.

VI. Conclusion

We presented an approach for grasping and reorienting large-scale boulders with an automated excavator, and evaluated its applicability in real-world experiments by assembling a dry stone structure. The manipulation approach itself is task agnostic and could be extended, besides of landscape construction, to industrial assembling or household and service robotics. To further improve construction process reliability, the assembly planner could already consider the current object pose for finding possible placement poses, to avoid time consuming reorientation and reduce the number of grasp attempts.

It is not realistic to achieve a perfect success rate on the first grasp attempt for a complex task like collision free assembly with irregularly shaped objects. Therefore, we plan to improve the autonomy of the grasp process with an automated strategy, inspired by how a human operator grasps irregular objects: slightly adapting the grasp during gripper closing instead of performing complete re-localization and re-grasping upon detection of stone slippage or dropping.

References

[1] K. H. Petersen, N. Napp, R. Stuart-Smith, D. Rus, and M. Kovac, “A review of collective robotic construction,” *Sci. Robot.*, vol. 4, no. 28, 2019.

[2] F. Gramazio and M. Kohler, *The Robotic Touch: How Robots Change Architecture*. Zürich, Switzerland: Park Books, 2014.

[3] P. Eversmann, “Robotic fabrication techniques for material of unknown geometry,” *Humanizing Digital Reality*. Berlin, Germany: Springer, 2018, pp. 311–322.

[4] K. Dörfler, T. Sandy, and M. Giftthaler, “Mobile robotic brickwork,” *Robotic Fabrication in Architecture, Art and Design 2016*, D. Reinhardt, R. Saunders, and J. Burry, Eds. Cham, Switzerland: Springer, 2016, pp. 204–217.

[5] R. L. Johns *et al.*, “Autonomous dry stone: On-site planning and assembly of stone walls with a robotic excavator,” *Construction Robot.*, vol. 4, no. 3, pp. 127–140, 2020.

[6] D. Jud, P. Leemann, S. Kerscher, and M. Hutter, “Autonomous free-form trenching using a walking excavator,” *IEEE Robot. Automat. Lett.*, vol. 4, no. 4, pp. 3208–3215, Oct. 2019.

[7] H. Dobashi *et al.*, “Robust grasping strategy for assembling parts in various shapes,” *Adv. Robot.*, vol. 28, no. 15, pp. 1005–1019, 2014.

[8] D. Prattichizzo and J. C. Trinkle, “Grasping,” in *Springer Handbook of Robotics*, B. Siciliano and O. Khatib, Eds. Springer International Publishing, 2016, ch. 38, pp. 955–988.

[9] 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, Apr. 2014.

[10] A. ten Pas and R. Platt, “Using geometry to detect grasp poses in 3D point clouds,” in *Proc. Robot. Res.*, 2018, pp. 307–324.

[11] D. Kappler, J. Bohg, and S. Schaal, “Leveraging big data for grasp planning,” in *Proc. IEEE Int. Conf. Robot. Automat.*, 2015, pp. 4304–4311.

[12] A. Herzog *et al.*, “Learning of grasp selection based on shape-templates,” *Auton. Robots*, vol. 36, no. 1-2, pp. 51–65, 2014.

[13] A. ten Pas, M. Gualtieri, K. Saenko, and R. Platt, “Grasp pose detection in point clouds,” *Int. J. Robot. Res.*, vol. 36, no. 13-14, pp. 1455–1473, 2017.

[14] C. Borst, M. Fischer, and G. Hirzinger, “Grasping the dice by dicing the grasp,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2003, pp. 3692–3697.

[15] W. Wan, H. Igawa, K. Harada, H. Onda, K. Nagata, and N. Yamanobe, “A regrasp planning component for object reorientation,” *Auton. Robots*, vol. 43, no. 5, pp. 1101–1115, 2019.

[16] T. Lozano-Pérez and L. P. Kaelbling, “A constraint-based method for solving sequential manipulation planning problems,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2014, pp. 3684–3691.

[17] Y. Hou, Z. Jia, and M. T. Mason, “Reorienting objects in 3D space using pivoting,” 2019, *arXiv:1912.02752*.

[18] N. Chavan-Dafle, M. T. Mason, H. Staab, G. Rossano, and A. Rodriguez, “A two-phase gripper to reorient and grasp,” in *Proc. IEEE Int. Conf. Automat. Sci. Eng.*, 2015, pp. 1249–1255.

[19] N. Chavan-Dafle *et al.*, “Extrinsic dexterity: In-hand manipulation with external forces,” in *Proc. IEEE Int. Conf. Robot. Automat.*, 2014, pp. 1578–1585.

[20] J. Shi, J. Z. Woodruff, P. B. Umbanhowar, and K. M. Lynch, “Dynamic in-hand sliding manipulation,” *IEEE Trans. Robot.*, vol. 33, no. 4, pp. 778–795, Aug. 2017.

[21] Y. Hou, Z. Jia, A. M. Johnson, and M. T. Mason, “Robust planar dynamic pivoting by regulating inertial and grip forces,” *Algorithmic Foundations of Robotics XII*. Berlin, Germany: Springer, 2020, pp. 464–479.

[22] R. Mascaro, M. Wermelinger, M. Hutter, and M. Chli, “Towards automating construction tasks: Large-scale object mapping, segmentation, and manipulation,” *J. Field Robot.*, to be pulished, doi: https://doi.org/10.1002/rob.22007.

[23] R. Dubé, A. Gawel, H. Sommer, J. Nieto, R. Siegwart, and C. Cadena, “An online multi-robot SLAM system for 3D LiDARs,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2017, pp. 1004–1011.

[24] G. Stoll, “Kostenrelevante Faktoren auf Trockenmauerbaustellen,” Accessed: Feb. 1, 2021. [Online]. Available: https://stonewalls.ch/sites/default/files/public/documents/Offer/ Kostenfaktoren.pdf