Received: 10 September 2020 | Revised: 17 March 2021 | Accepted: 30 April 2021

DOI: 10.1002/rob.22022

RESEARCH ARTICLE

WILEY

Autonomous robotic rock breaking using a real-time 3D visual perception system

Santeri Lampinen[1] | Longchuan Niu[1] | Lionel Hulttinen[1] | Jouni Niemi[2] | Jouni Mattila[1]

[1]Faculty of Engineering and Natural Sciences, Unit of Automation Technology and Mechanical Engineering, Tampere University, Tampere, Finland

[2]Rambooms Oy, Lahti, Finland

Correspondence
Santeri Lampinen, Faculty of Engineering and Natural Sciences, Unit of Automation Technology and Mechanical Engineering, Tampere University, Tampere, Finland.
Email: santeri.lampinen@tuni.fi

Abstract

Crushing of blasted ore is an essential phase in extraction of valuable minerals in mining industry. It is typically performed in multiple stages with each stage producing finer fragmentation. Performance and throughput of the first stage of crushing is highly dependent on the size distribution of the blasted ore. In the crushing plant, a metal grate prevents oversized boulders from getting into the crusher jaws, and a human-controlled hydraulic manipulator equipped with a rock hammer is required to break oversized boulders and ensure continuous material flow. This secondary breaking task is event-based in the sense that ore trucks deliver boulders at irregular intervals, thus requiring constant human supervision to ensure continuous material flow and prevent blockages. To automatize such breaking tasks, an intelligent robotic control system along with a visual perception system (VPS) is essential. In this manuscript, we propose an autonomous breaker system that includes a VPS capable of detecting multiple irregularly shaped rocks, a robotic control system featuring a decision-making mechanism for determining the breaking order when dealing with multiple rocks, and a comprehensive manipulator control system. We present a proof of concept for an autonomous robotic boulder breaking system, which consists of a stereo-camera-based VPS and an industrial rock-breaking manipulator robotized with our retrofitted system design. The experiments in this study were conducted in a real-world setup, and the results were evaluated based on the success rates of breaking. The experiments yielded an average success rate of 34% and a break pace of 3.3 attempts per minute.

KEYWORDS

computer vision, control, manipulators, mining, perception

Santeri Lampinen and Longchuan Niu should be considered as joint first authors.

1 | INTRODUCTION

Driven by safety and operational cost concerns, mining and construction automation systems have recently acquired foothold in various process phases of the mineral industry. However, many mineral processing tasks still involve unautomatized manual work that requires constant human supervision and intervention, which can act as a critical bottleneck for the process throughput.

One such task is secondary breaking, where controlled size reduction of mineral ore is achieved with heavy-duty manipulators equipped with hydraulic impact hammers. The mining industry extensively uses these types of rock breaker booms for size reduction of oversized boulders, which we will refer to as "rocks" in this paper. The economic justification for using such booms is to reduce process delays and ensure a steady flow of material, leading to minimal process downtime, maximized throughput, and increased productivity.

Rock breaker booms can be roughly divided into two categories based on their application. Small-scale breaker booms are used in mobile jaw crushers (see Figure 1a) to resolve material blockages, for example, for breaking oversized rocks entering the crusher cavity. In contrast, large-scale pedestal-mounted breaker booms (see Figure 1b,c) are mainly used in stationary grizzly applications, for example, in underground and surface mines, to process run-of-mine ore delivered by trucks. In grizzly applications, a steel grate is used as a screening medium to control the coarseness of the material entering an ore pass. In the event of buildup caused by oversized rocks that cannot pass through the openings of the grate structure, the rocks must be demolished into smaller particles using the hydraulic impact hammer.

Rock breaking systems require skilled and alert operators, since the interaction between the hammer and the rocks must be controlled with appropriate levels of force. Presently, rock breaking systems are largely operated via manual open-loop control of each individual joint, making their use inergonomic and unintuitive from the operator's point of view, thus increasing accident-proneness. Much of an operator's cognitive effort is dedicated to avoiding potentially dangerous and/or harmful situations, such as sudden loss of contact between the hammer and the rock, which might cause idle strokes of the hammer in the air–or worse, an unintended collision with the environment, which could deteriorate the hammer and shorten its lifespan (Sandvik Mining and Construction, 2016). Impact on the grizzly itself must also be avoided, as breaking it can lead into prolonged downtime in production. It has also been reported that nearly three out of four crane accidents are operator-induced (Lovgren, 2004), which is a strong argument for developing semi-autonomous solutions for rock breaking systems. With this in mind, it is worth noting that even human operators cannot achieve a 100% success rate in the breaking process, but will experience many failed attempts resulting from rocks being moved under the hammer during break attempts. Another strong argument for semiautonomous and autonomous systems is the fact that labor represents a major share of costs in underground mining operations (Hustrulid & Bullock, 2001). The fact that a large-scale underground mine can contain several crusher units further highlights the significance of automating this phase of the mining process.

To automate such breaking tasks in a harsh environment, the need for an intelligent robotic system with visual perception is evident. Human operators can easily distinguish between individual rocks on the grizzly and choose an ideal spot on the rock's surface to break it efficiently. However, real-time three-dimensional (3D) rock detection is challenging, as rocks come in arbitrary shapes, sizes, colors, and surface textures and do not follow any specific patterns. The high-precision control of the breaker boom presents another challenge, as the manipulators have been designed with manual operation in mind, and are thus typically equipped with slow control valves with highly nonlinear characteristics. In addition, a successful breaking process involves accurate pose estimation of the rock (the 3D position and 3D orientation of its major surface plane), precise calibration of the intrinsic and extrinsic camera parameters as well as the robotic manipulator itself, and a reliable decision-making mechanism that takes action autonomously after an oversized rock has been detected.

In this manuscript, we propose an autonomous robotic rock breaking system that utilizes the 3D object detection pipeline

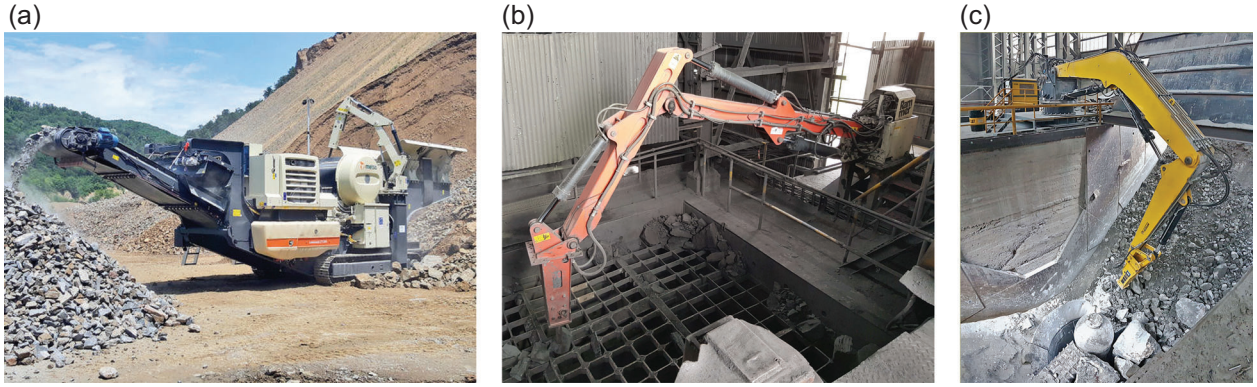

FIGURE 1 (a) A rock breaker boom on a Metso Locotrack mobile crusher, (b) a pedestal breaker boom in a grizzly application, and (c) a breaker boom at a gyratory crusher facility [Color figure can be viewed at wileyonlinelibrary.com]

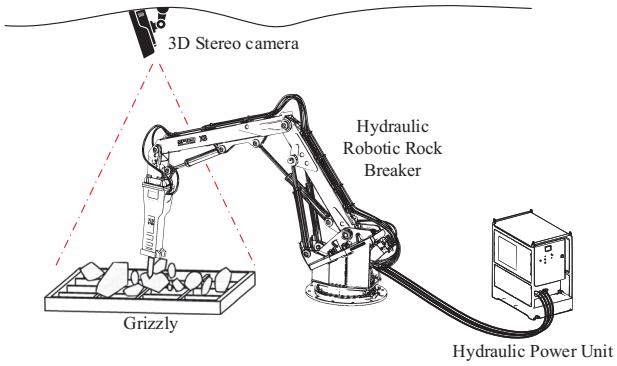

FIGURE 2 Conceptual illustration of the proposed autonomous rock breaker system [Color figure can be viewed at wileyonlinelibrary.com]

proposed in Niu et al. (2019) to automatically detect and localize rocks on the grizzly using a low-cost stereo camera. The rock positions are utilized by our real-time control system for which we have designed a robust decision-making mechanism along with a comprehensive manipulator controller, trajectory generator, and rock breaking control algorithm. Figure 2 illustrates the proposed system on a practical level. We present the measures conducted to precisely calibrate each subsystem, first separately and then together as a complete system. The outcome of this manuscript culminates on a field experiment of the system that acts as a technological proof of concept in a simplified environment.

1.1 | Literature review

Previous works concerning the automatization or modernization of rock breaking systems are few. The first reported attempts at autonomous vision-based rock breaking originates from 1998. Takahashi and Sano (1998) proposed an early image processing approach to detect rocks on the grizzly. The position of the rocks was obtained by complementing image data with a laser pointer mounted on the manipulator. Corke et al. (1998) proposed an actuated scanning laser rangefinder to identify and localize rocks. However, it was evaluated insufficient based on a concluded field test. In the field test, the rangefinder was positioned only slightly above the grizzly, and thus larger rocks on the grizzly blocked the view easily. The study discussed different visual sensing approaches as well, such as stereo vision, and proposed a concept of a semiautomated rock breaker. They identified several key requirements for an automated rock breaker system, such as a closed-loop controlled breaking boom, a 3D sensing system, an autonomous decision-making system, and a teleoperation system as a backup control method. They proposed a system that attempts to autonomously break rocks on the grizzly; if unable to complete the operation, it signals an operator to finish the job. With limited human intervention required, one operator could monitor several booms at the same time. The study was concluded, however, with a statement that the technology for such system is "many decades from reality."

The first teleoperated rock breaker was reported in Hubert et al. (2000). Designing a teleoperation system for the rock breakers was motivated by safety concerns. An underground mine in Indonesia was suffering from wet muck spills that placed the machine operators in danger. A communication system was designed to control the manipulators from a surface control room, but the machine operation was kept in open-loop manner. More recently, teleoperated rock breakers have been proposed by Duff et al. (2010), who demonstrated teleoperation over a distance of 1000 km over the internet. The breaker boom was also under closed-loop control, and the operator used resolved rate control to affect the velocity of the manipulator tip directly. Automatic deployment and parking of the manipulator was incorporated into the system with a mixed reality interface that combined computer-generated scene of the environment with reconstructed rocks on the grizzly. The 3D view from the grizzly was obtained using two stereo cameras. A more recent approach was reported in Boeing (2013) which discusses a system reportedly similar to the one presented by Duff, but it is accompanied by a collision avoidance system to prevent collisions with the environment.

Space exploration has also advanced the sophistication of vision based rock detection. In Fox et al. (2002), 2D camera images were combined with range data to detect larger rocks autonomously. A more recent study of the automatic detection of large rocks using a time-of-flight (TOF) camera, which is commonly used in the industry, was presented in McKinnon and Marshall (2014). The intended application was evaluating rock piles for excavation purposes. In Niu et al. (2018), a TOF camera was employed for rock detection on the grizzly, but the TOF camera's low resolution made it insufficient for the task. In Niu et al. (2019), a deep learning approach was presented in which the functionality of "you only look once" version 3 (YOLOv3), a state-of-the-art real-time object detection algorithm (Redmon & Farhadi, 2018), was extended from using 2D images to 3D point clouds for rock detection.

The notable lack of more recent reported automatized rock breaking applications indicate that there is further room for improvement and plenty of opportunities to apply visual perception and robotic control in rock breaking tasks, with the aim of making rock breaking systems safer, faster, and more efficient.

1.2 | Organization of the manuscript

The rest of this paper is organized as follow: Section 2 states the problem this manuscript aims to solve, along with the identified challenges and research objective. Section 3 presents the design of each subsystem of the proposed system. The section describes first the architecture on a high level, then in more detail about the visual perception and the control system design. Section 4 discusses the calibration of the manipulator and camera, as well as their integration into the same coordinate system. Section 5 presents the experiments with the proposed system and discusses about the obtained results. Section 6 discusses identified shortcomings of the

proposed system and suggests improvements to address these issues. Section 7 concludes the paper with a projection on future research potential in this area of study.

2 | PROBLEM STATEMENT

2.1 | Rock breaking—use case

Size reduction of blasted ore is an integral part of mineral extraction in mining. It is an essential process in the sense, that smaller ore pieces can be transferred more easily and also chemical/mechanical extraction methods can be applied to them. Size reduction of the vast majority of material is performed using a primary breaker (e.g., gyratory- or jaw-crusher), while oversized rocks, too big for the primary breaker, need to be broken with a secondary breaker.

Secondary breaking processes utilizing impact breaking can occur in multiple contexts, for example, directly at the blasting site using an excavator-mounted hydraulic hammer or with a special breaker boom at a gyratory crusher against the wall of the gyrator cone. In this study, our focus is on grizzly applications (see Figure 3), where a steel grating plate is used to prevent oversized rocks from getting into the primary crusher. The primary crushers are designed for a specific size reduction of the material flow, and overly coarse material can lead to material buildups or even material flow blockage, thus halting the entire operation.

The need for secondary breaking varies between mines and construction sites and depends on the material being processed. Even so, the need for secondary breaking is a symptom of imperfect blasting and problems in the blasting process. In ideal conditions, the blasting cycle is controlled to obtain material of a desired size (Zhang, 2016). When the process is well controlled, the need for secondary breaking is minimal.

Whenever oversized rocks are caught on the grizzly structure, the rock hammer is used to reduce their size. This temporarily halts material flow; for example, an ore truck must stop feeding material to a silo until the breaker boom operator breaks the oversized rocks into smaller pieces that can pass through the grizzly. If the boom cannot execute its task in a limited time frame, the rock is pushed away from the grizzly for later processing and the arm returns to its resting position. The material that cannot pass through the openings of the screening medium should be broken with a hydraulic hammer. This process is referred to as screening, which is an essential step in crushing unprocessed run-of-mine ore and turning it into a finer substance suitable for further treatment (Metso Mining and Construction, 2015).

The actual use case studied here can be described as the process of breaking an oversized rock caught in the grizzly. Additional use cases in grizzly applications are raking with the boom to break and prevent blockages, and reorienting hard-to-break rocks for easier breaking. The current study is limited to the breaking process. The studied use case can be described on a high-level with the following steps:

(1) The boom is driven from a rest position to a standby position beside the grizzly, with its hammer kept at a 90-degree angle relative to the grizzly.
(2) A 3D visual perception system (VPS) detects and localizes oversized rocks on the grizzly and passes the information on to the main control system.
(3) The main control system determines the shortest rock-to-rock trajectory from the information provided by the VPS, employing a lower level control system to break each rock.
(4) The path planner receives the target rock coordinates from the high-level controller and generates a trajectory from the manipulator's current position to a position above the target rock.
(5) An approach motion toward the target rock is performed while maintaining the desired tool orientation.
(6) When target coordinates are reached, the boom maintains pressure against the rock and switches the rock hammer on.
(7) After the rock has been broken, the boom shall rise up to a safe transition height and wait for the next target from the high-level control system.

FIGURE 3 Rambooms X88-540R breaker boom at the field test site at Tampere University [Color figure can be viewed at wileyonlinelibrary.com]

(8) After clearing the rocks, the boom returns to the standby position to wait while the VPS inspects the work and identifies remaining rocks on the grizzly.

A critical issue in rock breaking is to make contact with the rock in a controlled manner and with sufficient force against the rock. In the case of grizzly applications, tool alignment is an important issue, as the supportive force from the grizzly points upward and there is not necessarily anything holding the rock in place in the horizontal plane. In these scenarios, roughly a 90-degree angle relative to the grizzly is the most suitable (see Figure 3). An incorrect breaking angle may cause excess wear and stress to the manipulator or the rock can slip away under the hammer. Situations in which a hydraulic cylinder is at its mechanical stroke limit during hammer operation must be avoided. Given all these concerns, significant attention and effort is necessary to avoid dangerous situations and achieve a good contact with the rock.

2.2 | Challenges

To implement an autonomous system for the rock breaking process, we have identified four distinct main challenges we will need to consider and solve. The challenges are related to: (1) The visual perception, (2) the autonomous operation strategies, (3) the high-precision manipulator control and stable contact control, and (4) system calibration and integration.

To achieve autonomy in the rock breaking process, it is crucial for the robot to properly understand the scene. However, detecting each individual rock in a cluttered and dynamic scene is a highly complex activity, as rocks cannot be characterized by any particular feature. They may possess a variety of colors, unique surface textures and arbitrary shapes and sizes. Despite these challenges, the VPS should operate robustly under dynamic outdoor weather conditions being able to accurately detect all rocks in the grizzly. The detection must also include rocks partially occluded by overlapping rocks or the manipulator arm. The VPS should propose a suitable breaking position based on the surface of each rock.

Next, we need a robust and efficient strategy for autonomous operation. The decision-making process should consider the shortest trajectories between rocks and have the ability to govern manipulator movement sequences. To properly make decisions, perception information from the vision system must be assessed and cataloged. In addition, the system should discern valid rock positions and discard any invalid positions received from the perception system. Locations may be considered invalid for rocks below the grizzly and rocks outside the grizzly.

Building the control system for the robotic manipulator is another challenge that requires sophisticated and rigorous solutions. As the manipulator is not retrofitted with fast servo valves and has a slow natural frequency, its precise control requires thorough consideration. Other constraints, such as tool orientation and flow rate limitations need consideration as well.

For contact control, we assume the accuracy of the manipulator's tool center point (TCP), which is the tip of the hammer, to stay at all times within the initial requirement of 150 mm from the target position. As rocks are typically much larger than this and the mesh size in our testing grizzly is 400×400 mm, this accuracy requirement is reasonable. Based on our preliminary experiments, the most challenging task is making contact with rock surfaces so that they do not slide under the hammer or tip over. Since blasted boulders come in arbitrary shapes and sizes with sharp edges, they end up laying on the grizzly randomly. As a consequence, the following two main challenges apply to rock breaking: first, a boulder or multiple piled boulders may be poorly balanced on the grizzly and thus cannot support the required hammer tip loading force without rotating into new orientations, slipping away from the applied contact force and thus failing to break. Second, if a boulder has inadequate flat surface area for firm hammer contact force, the hammer tip may slide along the rock without breaking it.

Uncertainty about subsystem-level accuracies is also a challenge in estimating the final system performance and accuracy. Individual subsystem calibration for the robotic manipulator and stereo camera is required to estimate the accuracy of the final autonomous system. Causes of uncertainty about the accuracy of the final system can be the precision of rock detection model, the accuracy of the intrinsic and extrinsic camera calibration, kinematic parameters of the manipulator used to calculate TCP position, and control system accuracy.

The most important challenge, however, is integrating all the distinct subsystems together with their respective safety functions. Responsibilities and communication between subsystems can be vague and multifaceted, and managing their complexity is critical.

2.3 | Research objective

The primary objective of this manuscript is to demonstrate a proof of concept for an autonomous hydraulic breaker boom system. The aim of this manuscript is not to showcase a finished product, but rather to demonstrate the feasibility of the concept. This should be noted when evaluating the experimental results and required hardware.

The major function of the robotic VPS is twofold: first, achieving a fast and robust 3D rock detection mechanism regardless of rock shapes and sizes in overlapping scenarios, and second, providing reliable positions for the manipulator to break rocks. The objective of studying visual perception systems is to assess their effectiveness in detecting objects with unpredictable features for heavy duty manipulator applications.

From the control system point of view, the objectives can be categorized as the desired control accuracy of the manipulator and the desired behavior of the autonomous functions and safety features. Given the size of the rocks being broken, the absolute accuracy of the control system should be within 150 mm, which an

interview with a domain expert substantiated. Manipulator limitations, such as the size of the control valves, that define maximum velocity for each actuator, maximum volumetric flow rate of the hydraulic supply unit that limits maximum endpoint velocity, and the reachable workspace, must be taken in account when designing the control system.

Our goal is to make the manipulator independently decide an intelligent rock breaking order based on the data provided by the VPS, generate trajectory between each rock, execute the trajectory in the breaking process. While the chance of successful breaks will not be high initially, we will also endeavor to make the system detect rocks from the grizzly during operation and adjust its plan in real time. Safety functions built into the control system prevent impact to the grizzly during the breaking process to avoid damage to the hammer and premature component failure.

3 | SYSTEM DESIGN

3.1 | High-level architecture

The proposed system comprises of three distinct parts: the instrumented hydraulic breaker boom and its hydraulic power unit, the VPS and the real-time control system that governs decision-making, the manipulator control system, and all measurement data. The complete system is depicted in the high-level architecture diagram in Figure 4.

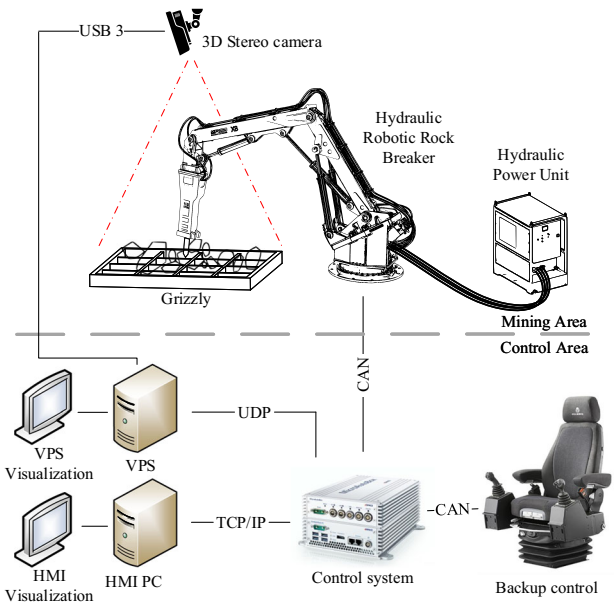

FIGURE 4 High-level architecture of the proposed system. For clarity, the site cameras surrounding the crushing site and their visualization computer has been left out. HMI, human machine interface; TCP, tool center point; UDP, user datagram protocol; VPS, visual perception system [Color figure can be viewed at wileyonlinelibrary.com]

3.1.1 | Hardware architecture

The hydraulic breaker boom used in this study was the commercial Rambooms X88-540R manipulator equipped with a Rammer 2577 hydraulic impact hammer. The breaker boom weighs in total over 10,000 kg and has a horizontal reach of 5.4 m with the breaker in vertical orientation. The coordinate frame assignment of the manipulator along with the joint naming convention is shown in Figure 5. The link lengths a_2, a_3, and a_4 in Figure 5 are all roughly 3 m. The size of the grizzly (see Figure 3) is 2.6 m × 4.0 m. The manipulator was retrofitted with Siko WV58MR 14-bit absolute rotary encoders for joint angle measurements. The sensor data and the valve controls were transmitted to and from the real-time control system via CAN-bus interfaces.

The 3D VPS consists of a ZED stereo camera and a Linux PC. The stereo camera is mounted on a pole approximately 5 m above the workspace such that the grizzly is centered in the camera's field of view. The 3D VPS is connected to the real-time control system through a user datagram protocol (UDP) interface.

The real-time control system was run on a dSpace MicroAutoBox 2 real-time controller, where all control computations and decision-making logic were performed. The interface for the real-time controller was created using the dSpace ControlDesk software on a separate human machine interface (HMI) PC.

3.1.2 | Software architecture

The software architecture is divided into two parts based on the hardware architecture; The VPS running on a linux PC and the control system running on the dSpace real-time controller. The VPS is responsible for perceiving rocks on the grizzly, using the data from the stereo camera to detect and localize rocks and estimate the pose of the major surface plane near the highest point of the rocks. The

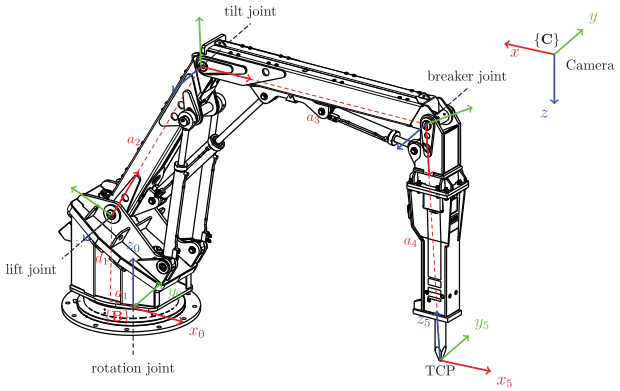

FIGURE 5 Coordinate frame assignment for the breaker boom. Frame {B} denotes the base coordinate frame of the manipulator, while frame {C} denotes the coordinate frame of the stereo camera. Joint naming convention is also depicted on the figure and the TCP is marked. TCP, tool center point [Color figure can be viewed at wileyonlinelibrary.com]

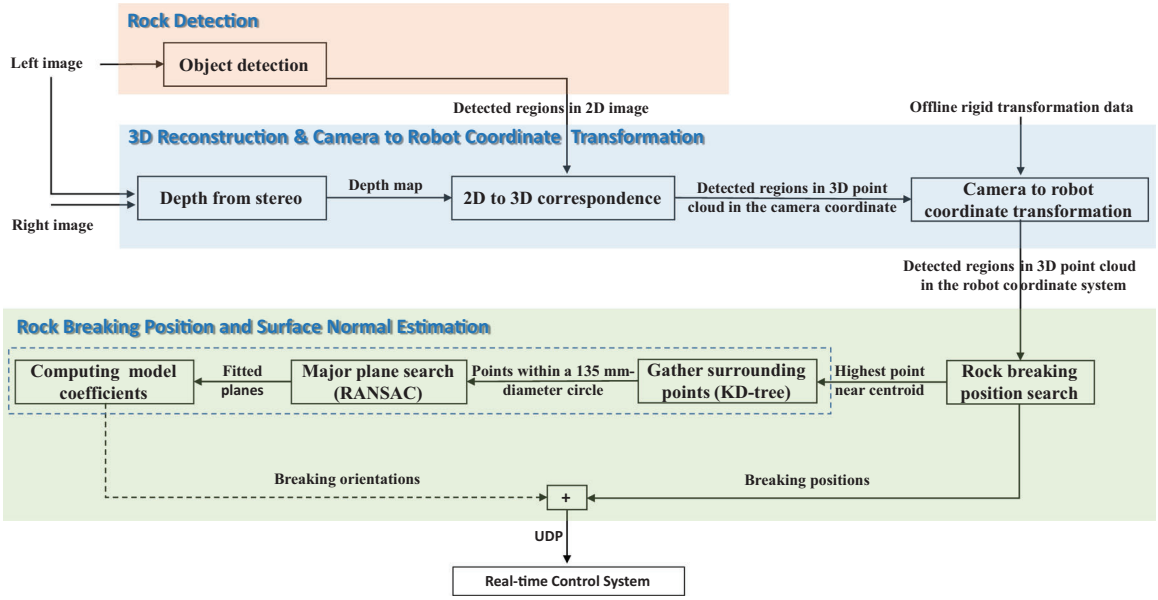

FIGURE 6 Workflow architecture of the 3D visual perception system [Color figure can be viewed at wileyonlinelibrary.com]

operation of the VPS is described in detail in Section 3.2. The real-time control system is responsible for decision-making related to break order logic, controlling the movements of the manipulator, and managing the safety functions. The operation of the control system is described in Section 3.3.

3.2 | Visual perception system

A high-level architecture for the workflow of the 3D VPS is illustrated in Figure 6, which consists of three stages: rock detection, 3D reconstruction and camera to robot coordinate transformation, and position and orientation estimation for rock breaking. At the first stage, the object detection module processes the left images of the ZED stereo camera and extracts the detected rocks as 2D regions. At the second stage, the detected 2D regions are reconstructed into 3D point clouds in the camera coordinate system with the aid of calibrated intrinsic camera parameters and the depth map. Then, the detected rock regions in 3D point clouds are transformed into the manipulator's coordinate system. At the last stage, the positions required to break each rock are determined by searching the highest point near the centroid of each region. The surface normals of each rock are estimated (in the dashed area in Figure 6) using KD-tree and RANSAC.

3.2.1 | 3D sensing modalities

Common 3D visual perception sensors are Lidar sensors, TOF cameras, and stereo cameras. The 3D sensor selected for visual perception must account for the aforementioned design challenges. At the boundary distance of 5 m to the grizzly, the mesh (400 × 400 mm) and rocks of similar size may appear small in the field of view. The empirical study with a TOF sensor (Niu et al., 2018) implies that spatial resolution and the amount of available information from a scene are decisive factors in accurate rock detection.

Lidar is gaining popularity across industries. However, compared to high-resolution images, Lidar point clouds are unstructured; as such, generic convolutional neural network (CNN) are not well suited to process them directly (Qi et al., 2019). In addition, relatively sparse Lidar point clouds can be inadequate in assessing the details of a scene where a pile of small irregularly shaped rocks are overlapping each other. In fact, current 3D object detection methods in Lidar applications have been targeted for use with spatially independent objects (Al Hakim, 2018; He et al., 2020; Ku et al., 2018; Liang et al., 2019, 2018; Qi et al., 2018; Yang et al., 2018; Zhao et al., 2019; Zhou & Tuzel, 2018). In contrast, an industry-ready stereo camera provides both high resolution images and dense point clouds. Its images contain rich texture information which is a useful cue for discriminating objects from the background. Therefore, we adopted a stereo camera in this study.

A camera setup can be classified as eye-in-hand or eye-to-hand. The eye-in-hand configuration is known as a close-range camera, which is rigidly attached to a robot's end effector. For rock breaking, this setup requires sustainable solutions to the following challenges: (1) involvement of robot and eye-in-hand calibration errors, (2) susceptibility to heavy vibrations, and (3) fragile lenses in close proximity to hazardous rock breaking operations. In light of these challenges, we considered eye-to-hand configuration, in which a compact ZED stereo camera is mounted on a pole 5 m above the workspace.

3.2.2 | Object detection

Three-dimensional object detection is one of the most prominent research areas of visual perception that serves as base for autonomous robotic tasks. As one of the main challenges in autonomous rock breaking, rock detection requires a deep understanding of the contexts of a scene. Background removal with semantic segmentation is inefficient, as this task requires every rock to be made visually distinct from one another in a cluttered and dynamic scene.

In recent years, deep learning frameworks have been available to computer vision applications to assist learning of deep and high-level features. The substantial improvements to object detection have mostly been applied to 2D images rather than 3D point clouds. Generally, 2D convolution-based detection approaches are more sophisticated than 3D ones in industrial deployment. Among a number of 2D object detection architectures, region-based methods like region-based convolutional neural networks (R-CNN) (Girshick et al., 2014), Fast R-CNN (Girshick, 2015), and Faster R-CNN (Ren et al., 2015) are accurate for detecting multiple objects in an image. However, their rather complex architectures and relatively low detection speeds are not sufficient for our purposes. In addition, the potential source of errors is high due to their complexity.

Mentioned briefly in the literature review, the object detection algorithm YOLOv3 prioritizes both recognition and speed. It is an improved version of the initial release of YOLO (Redmon et al., 2016), that used a new approach to object detection. Instead of repurposing classifiers to perform detection, YOLO uses a single neural network to predict bounding boxes and class probabilities from a full image. The third version, YOLOv3, is the result of incremental updates (Redmon & Farhadi, 2017, 2018), and it achieves high precision and high speeds on benchmark data sets; as such, the infrastructure of our deep learning network for object detection is based on YOLOv3.

The next step in deep learning is gathering data, the quality and quantity of which will determine the performance of the model. Our rock image data set initially contained 4733 distinct images[1] collected from the field test site (see Figure 3), where the amount of rocks varied between 1 and 15. These images were taken in September and October of 2018. The image data set contains images taken under sunny daylight condition. Images exhibiting other seasonal and weather-based conditions, such as rain, snow, and fog, are missing.

To emulate these missing weather conditions, synthetic data via data augmentation can be used to bridge the experiment-reality gap. Generating realistic environmental variant data can be achieved using OpenCV and NumPy in Python. Besides different weather conditions, also dynamic lighting can cause challenges for the stereo camera and the model. For example, rock edges may become indistinguishable under bright lighting conditions. With this in mind, our data augmentation process involves generating portions of brighter images for labeling. This way the original data set was expanded to a total of 23,850 images. More training data from situations the model cannot cope with might be used to further improve it. Such conditions may include for example, low and bright lighting, and partly shaded rocks.

Our image data set contains only one class: the "rock" class. The data set was split into three parts: 70% images for training the model, 20% images for validation, and 10% images for testing. The training was conducted on YOLOv3's darknet-53 architecture (Redmon, 2018) on an Ubuntu 16.04 Linux PC with a NVIDIA Quadro P5000 graphics card. The training step used our training set to incrementally improve the model's ability to make inferences, while each epoch updated the weights of the model. The training converged at an average loss of 0.12 with a batch size of 64 and a learning rate of 0.001. An evaluation experiment given in Figure 7 illustrates the results of the model inference after training. It also points to the improvement gained through data augmentation.

To further evaluate the performance of our model, we used the average precision (AP) metric to compute precision and recall by Equation (1), where TP denotes the number of true positives, FP the number of false positives, and FN the number of false negatives.

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN}. \tag{1}$$

Table 1 shows the test with the AP metric, where AP_{50} and AP_{75} denote the average precision computed at an intersection over union (IOU) threshold of 0.5 and 0.75, respectively. An average detection speed of 85 ms per frame was achieved during testing.

3.2.3 | Establishing 2D to 3D correspondence

Estimating scene geometry from a pair of pinhole cameras is often referred to as depth-from-stereo. For ease of setup, we employed a ZED stereo camera. From the left and right images of a stereo camera, its depth map can be generated with a rectification-based stereo-matching method (Scharstein & Szeliski, 2002) or plane-sweeping method (Smirnov et al., 2015). A depth map is an image representing the depth information of the scene associated with the corresponding left and right images of the stereo camera.

With the left image and associated depth map, a 3D point cloud of the scene can be reconstructed with the camera's intrinsic parameters. As illustrated in Figure 8, this 3D reconstruction process is known as triangulation, which can be applied to detected regions in an image to generate detected regions in a 3D point cloud. Processing a 3D point cloud of only the detected regions instead of the whole image decreases the associated computational burden.

3.2.4 | Determining the breaking position for each rock

The detected rocks are represented as rectangular regions in a 2D image. The position of the geometrical center of each detected rectangular region is used to describe each rock position in the image

[1] https://github.com/epoc88/SecondaryBreakingDataset

TABLE 1 Average detection rates of our model

	AP_{50} (%)	AP_{75} (%)
Proposed method	99.00	97.61

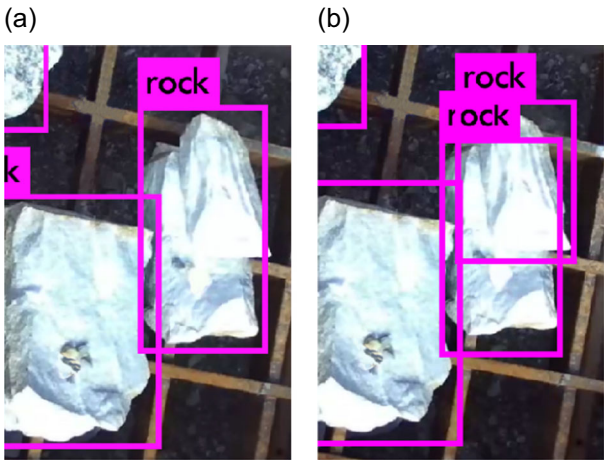

FIGURE 7 Compared detection results following data augmentation. The scenario depicts a smaller rock on top of a bigger rock under overexposed lighting conditions. (a) Original model (Niu et al., 2019) and (b) improved model [Color figure can be viewed at wileyonlinelibrary.com]

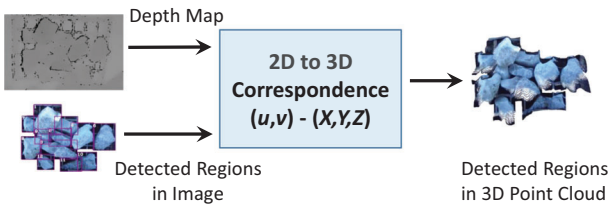

FIGURE 8 This figure illustrates the process of obtaining 3D point cloud of detected regions from stereoscopic imagery. To each pixel (*u*,*v*) in a detected region, there is corresponding depth information in the depth map. Combining these two sources for each detected region, we acquire a 3D point cloud representation of the detected objects in the camera's coordinate system [Color figure can be viewed at wileyonlinelibrary.com]

coordinate system, which can be transformed into the robot coordinate system by using the calibrated intrinsic and extrinsic camera parameters, see Figure 17.

The geometrical centroid of the identified rectangle itself is not necessarily an ideal breaking position due to fact that the detection algorithm does not take the shape of the rock into account. A better alternative for the breaking position can be obtained instead by searching for the highest point near the identified centroid. An examplatory case is depicted in Figure 9. The centroid position as it appears to the camera is not an ideal breaking position, and the attempt would fail with a high likelihood due to a probable rock movement.

Based on our preliminary field tests, the highest point near the centroid of a rock typically yields the highest likelihood for successful breaking. Thus, we limit the search area to a concentric rectangle quarter the size of the detected region. Breaking positions outside of the search area are discarded, in view of the fact that the likelihood of the manipulator slipping or the rock moving increases when the breaking position is located near the edges of the rock.

3.2.5 | Estimating the rock surface normal

At the time of breaking, the tip of the manipulator's blunt tool is in contact with the rock's breaking position. The contacted area must be within roughly 70 mm of the highest point, as the diameter of the manipulator's blunt tool is 135 mm. To transfer the energy of the impact hammer to the rocks most efficiently, the orientation of the hammer must be perpendicular to the surface of the rocks. To achieve this, the orientation of the rock surface must be estimated. This process is divided into three steps:

(a) *Gather surrounding points*: A KD-Tree algorithm (Bentley, 1975) is used to search for points contained within a sphere with the same diameter as of the blunt tool and centroid at the breaking position. The search yields a cluster of points in the form of circular areas at each rock surface. For a visualization, see the points colored in blue in Figure 10.

(b) *Major plane search*: This step analyzes every cluster of points and carries out plane fitting with a RANSAC algorithm (Fischler &

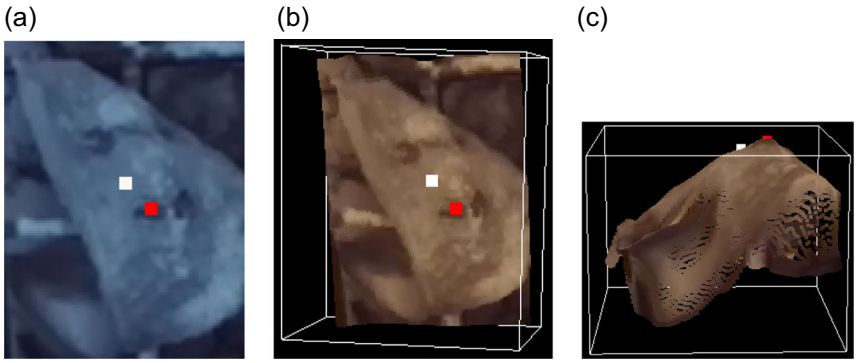

FIGURE 9 A detected rock region in a 2D image and a 3D point cloud. The white and red dots in the figure indicate the centroid of the detected region and the highest point within a quarter of the size of the detected region, respectively. (a) The rock in 2D image, (b) a point cloud from above, and (c) a point cloud from the side [Color figure can be viewed at wileyonlinelibrary.com]

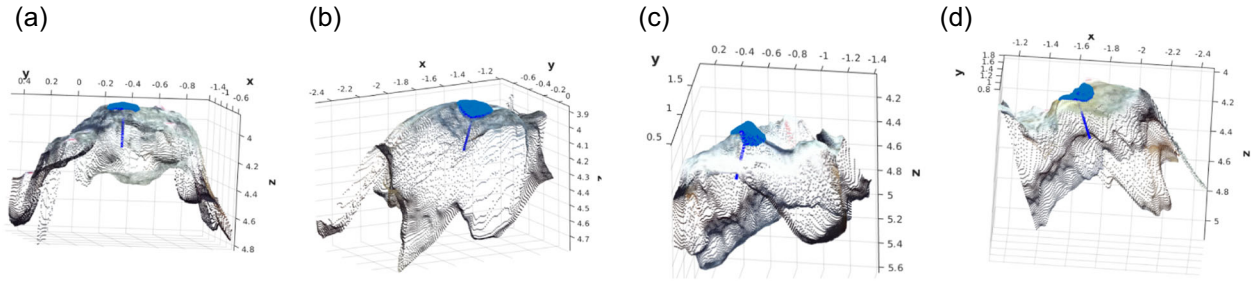

FIGURE 10 Some examples of estimated surface normals. The blue clusters are the rock surface points nearby each breaking position, and the blue arrows indicate the estimated surface normals. (a) Rock 1, (b) rock 2, (c) rock 3, and (d) rock 4 [Color figure can be viewed at wileyonlinelibrary.com]

Bolles, 1981). The algorithm randomly takes three points in the cluster to establish a plane. Points lying close to the plane are considered the consensus set for the plane. This process repeats until all the planes in the cluster are found; the plane with the largest consensus set is accepted as the fitted plane.

(c) *Computing model coefficients*: Finally, the model coefficients of each plane are computed to obtain the corresponding normal vectors of the plane. An example of the results of this process are shown in Figure 10.

In the conducted experiments, the 90 degree orientation of the hammer to the surface at the point of contact was not applied. Instead 90 degrees relatively to the grizzly was used.

3.3 | Control system

The control system design is depicted at a general level in Figure 11. The control system can be divided into four distinct subsystems with specific tasks. The breaking order logic and path optimization initializes the pipeline, working at a high level to determine rock breaking order. The second highest level subsystem is the high-level manipulator controller, a state machine that dictates the operation of the manipulator. The third level consists of the trajectory generator and is closely interconnected with the inverse kinematics controller and the flow rate limitation algorithm. The lowest level controller is used for the actual manipulator control, which uses desired joint angles and velocities as well as the operational state of the hydraulic hammer.

3.3.1 | Break order logic

The break order logic subsystem is devised around the idea that the manipulator might be blocking the camera's view, making it inevitable that the logic would store previous rock locations sent by the VPS. The path optimization should minimize movement between rocks. The optimal trajectory for breaking each rock in a sequence could be obtained by finding a solution to the classical traveling salesman problem, in which a traveling salesman seeks to find the shortest path that visits each city exactly once and return to the origin. To limit the complexity of our solution, we opted for a simple heuristic nearest neighbor approach with some additional constraints. The developed approach is showcased in the high-level

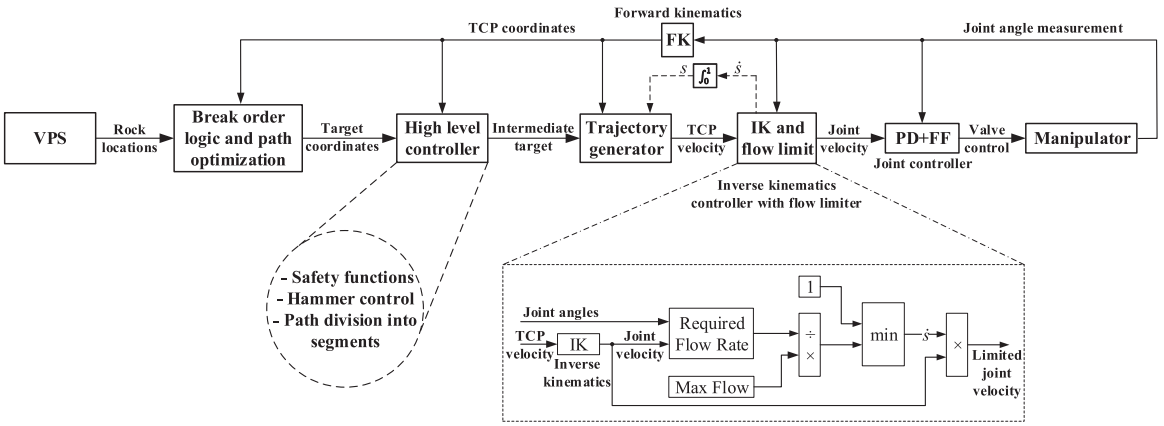

FIGURE 11 General block diagram of the proposed control system for autonomous operation. The VPS in the first block on the left is described in more detail in Figure 6

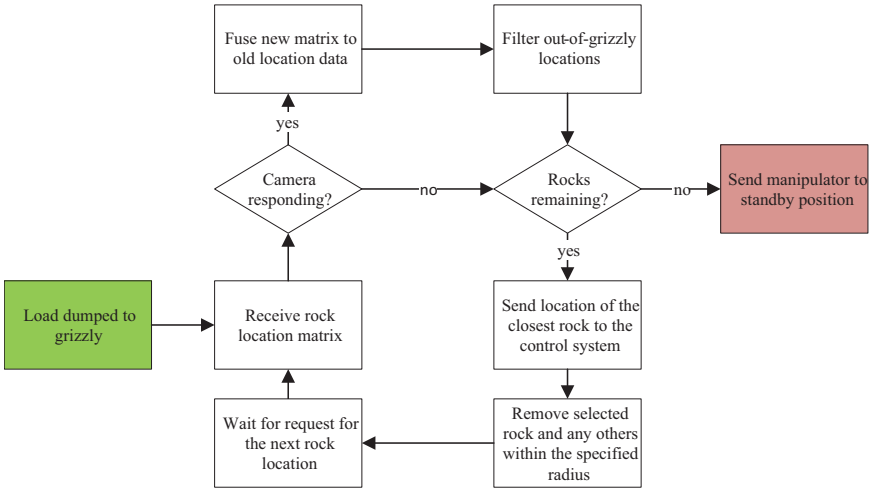

FIGURE 12 Breaking order logic pipeline. The start of the process is marked with green color and the end with red. The loop in the middle is continued until no rock are remaining on the grizzly [Color figure can be viewed at wileyonlinelibrary.com]

diagram in Figure 12. The pipeline can be described by the following steps:

(1) The cycle starts when a load of rocks is dumped on the grizzly, and the system receives command to begin operation. In our experiments, the cycle was started manually.
(2) First, rock positions from the VPS are obtained via UDP messages. If the vision system does not respond within a specific time frame (e.g., camera view is blocked by manipulator), the next target is determined using existing data. At the first round, the manipulator is at its standby position and not blocking the view.
(3) The received data is then fused into the existing location data. This step is omitted in the first round. The data fusion is performed by calculating the Euclidean norm between each rock from the old and the new data set. If the norm between a rock from the old and the new data set is less than or equal to 0.1 m, the rocks are assumed to be the same, and the old position for that particular rock is updated to correspond to the newly obtained information. If the norm is greater for all rocks in the old data set, the rock is assumed to be new, and it is added to the data set. The algorithm is described using pseudo code in Algorithm 1.
(4) Rocks that are out of the grizzly area and possible misidentified points, for example, due to a manipulator blocking the view, are filtered out from the data set.
(5) After filtering, the rock closest to the TCP is selected as the next target to be broken.
(6) The data set cleanup is followed then. The rock selected for breaking is removed first. Then, based on Remark 1, rocks that are within 0.5 m of the selected rock are also removed, as they may be shifted by the break attempt. Aging of data could also be utilized for more robust operation (e.g., rocks that have not been detected by the vision system for a set number of rounds can be assumed invalid).
(7) The system is then suspended until a request for a new target is received, that is, the manipulator has finished the break attempt of the last target.
(8) After receiving request for the next target, the systems resumes operation from Step 2.
(9) After no more rocks are found by the VPS nor any are remaining in the data set, the system informs the high-level manipulator controller and the boom is driven to its standby position.

Algorithm 1 Data fusion algorithm

Input: Stored position matrix $\mathbf{P}_{\text{memory}}$, New position matrix $\mathbf{P}_{\text{camera}}$

Output: Data sets fused into $\mathbf{P}_{\text{memory}}$

for each $p_{\text{new}} \in \mathbf{P}_{\text{camera}}$ **do**
  newRock ← True
  for each $p \in \mathbf{P}_{\text{memory}}$ **do**
  if $|| p_{\text{new}} - p ||_2 \leq 0.1$ **then**
    newRock ← False
    $p \leftarrow p_{\text{new}}$
  end if
  end for
  if newRock **then**
  $\mathbf{P}_{\text{memory}} \leftarrow \begin{matrix}\mathbf{P}_{\text{memory}} \\ p_{\text{new}}\end{matrix}$
  end if
end for

Remark 1. Based on our observations from preliminary experiments, an attempt to break a specific rock will not affect rocks that are not in the immediate proximity of the rock being broken. A 0.5 m radius is sufficient margin beyond which rocks will not be shifted by the broken rock. Due to the vibrations during the hydraulic hammer operation, rocks might move slightly farther away than expected, but the total movement of the rocks remains minor. However, any rocks inside the set radius are likely to move considerably. This has been tested only in situations, where the

rocks are resting on the grizzly in a single layer, and may not be valid in other situations.

3.3.2 | High-level manipulator controller

The high-level manipulator controller is an event-triggered state-machine that defines different operating modes of the test manipulator. In this application, three operational states are defined as follows: automatic unfolding, automatic folding, and autonomous rock breaking. In its nonoperational state, the main motion controller of the manipulator is disabled for safety reasons. The nonoperational state is defined as the default initialization state.

The automatic motion states move the manipulator from its current position to specific predetermined positions within the workspace of the manipulator. These positions are called standby position and resting position, respectively. The boom is driven to these positions through the following steps: first, the TCP is driven to a specified transition height. Then, the target is set to the *XY* coordinates of the prespecified position. Finally, the TCP is driven to the final target position. Note that the last movement of the TCP is only vertical.

The autonomous operation pipeline follows a specific pattern. First, the system requests target position from the break order logic subsystem. After a new target is obtained, the manipulator is raised to the transition height, if not already at that height, after which it is driven above the target rock. The approach move is triggered next, and this phase is linked to the breaking sequence. The approach move is executed so that the manipulator is set to drive 50 mm below the rock's surface to load the internal spring of the hydraulic hammer. After reaching the rock surface, all joints but the lift joint are locked to prevent the TCP from slipping away from the rock. The lift joint is used to maintain pressure against the rock. The hydraulic impact hammer is then engaged and kept on for 5 s or until the tip of the manipulator has entered a virtual safety zone, which is set 50 mm above the grizzly as a collision-avoidance measure. After the break attempt, the manipulator is driven back up to the transition height. The sequence is then repeated from the first step.

Remark 2. After the first experiments, the autonomous breaking sequence was revised so that after every third attempt, the manipulator moves aside to the standby position to give the stereo camera a clear view of the grizzly.

Remark 3. The modular system design allows for rapid testing of different approaches for breaking rocks. Contact and external force estimation are particularly interesting research topics, here omitted, that may notably increase the success rate of the break attempts. Impedance control has been proposed as one possible solution to achieve the required compliant behavior (Hulttinen, 2017; Koivumäki & Mattila, 2017; Tafazoli et al., 2002). At this stage, a strategy for approaching the rocks without them slipping and moving away could be devised. Learning from demonstrations is another interesting and seemingly promising approach for instructing robots on contact control with teleoperated demonstrations from a human operator (Havoutis & Calinon, 2019; Suomalainen et al., 2018).

3.3.3 | Trajectory generation and inverse kinematics

The trajectory generator for the manipulator is designed to generate trajectories from the current position of the manipulator's TCP to the target coordinates. Trajectories are created in a cylindrical coordinate system to minimize unnecessary actuator movements. The trajectory generator first converts the start and end coordinates to the cylindrical coordinate system, respectively. Then, quintic rest-to-rest trajectories are created between the two points using

$$x(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3 + a_4 t^4 + a_5 t^5, \tag{2}$$

where x contains an individual point-to-point trajectory, and coefficients $a_i \in \mathbb{R}$ are obtained using

$$\begin{bmatrix} 1 & t_0 & t_0^2 & t_0^3 & t_0^4 & t_0^5 \\ 0 & 1 & 2t_0 & 3t_0^2 & 4t_0^3 & 5t_0^4 \\ 0 & 0 & 2 & 6t_0^2 & 4t_0^3 & 20t_0^3 \\ 1 & t_f & t_f^2 & t_f^3 & t_f^4 & t_f^5 \\ 0 & 1 & 2t_f & 3t_f^2 & 4t_f^3 & 5t_f^4 \\ 0 & 0 & 2 & 6t_f & 12t_f^2 & 20t_f^3 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix} = \begin{bmatrix} x_0 \\ \dot{x}_0 \\ \ddot{x}_0 \\ x_f \\ \dot{x}_f \\ \ddot{x}_f \end{bmatrix}, \tag{3}$$

where t_0 is time at the beginning and t_f is time at the end. x_0, $\dot{x}_0$, and $\ddot{x}_0$ denote the initial position, velocity and acceleration, respectively, whereas x_f, $\dot{x}_f$, and $\ddot{x}_f$ define the final position, velocity and acceleration, respectively (Jazar, 2010).

The trajectory generator provides the position and velocity along the path in Cartesian coordinates, but those must be transformed into joint space for the joint controller. Let $\mathbf{v} \in \mathbb{R}^3$ denote the desired velocity of the manipulator in robot coordinates. For a redundant four-joint manipulator, the required joint velocities can be identified using a pseudo-inverse of the Jacobian matrix, which is defined as

$$\mathbf{J}^\dagger = \mathbf{W}^{-1}\mathbf{J}^T(\mathbf{J}\mathbf{W}^{-1}\mathbf{J}^T)^{-1}, \tag{4}$$

where $\mathbf{J}^\dagger \in \mathbb{R}^{4\times4}$ is the Jacobian pseudo-inverse, $\mathbf{W} \in \mathbb{R}^{4\times4}$ is a symmetric positive definite weighing matrix, and $\mathbf{J} \in \mathbb{R}^{3\times4}$ is the non-invertible Jacobian matrix (Sciavicco et al., 2000). The weight matrix W is updated dynamically based on the joint configuration and the direction the joints are moving to prevent any actuator from reaching its mechanical stroke limits. Near the mechanical stroke limits, the weight of the corresponding actuator increase and thus prevent it from reaching mechanical limits. For more detailed description see (Lampinen et al., 2020).

The redundancy of the manipulator is utilized to control the angle of the hammer with respect to the ground. To change the pose of the manipulator without moving the TCP, we use the null space of the Jacobian matrix. The null space $\mathcal{N}(\mathbf{J})$ is obtained using

$$\mathcal{N}(\mathbf{J}) = I - \mathbf{J}^{\dagger}\mathbf{J}. \tag{5}$$

The joint velocities with null space control are finally calculated as

$$\dot{\mathbf{q}} = \mathbf{J}^{\dagger}\mathbf{v} + \mathcal{N}(\mathbf{J})\dot{\mathbf{q}}_0, \tag{6}$$

where $\dot{\mathbf{q}}_0 \in \mathbb{R}^4$ is the joint control term that changes the pose of the manipulator without affecting the position or velocity of the end-effector, while $\dot{\mathbf{q}} \in \mathbb{R}^4$ denotes the joint velocities corresponding to the Cartesian velocity $\mathbf{v} \in \mathbb{R}^3$.

3.3.4 | Flow-rate limitation

Hydraulic systems are characterized by many nonlinearities and constraints specific to hydraulics. An important restriction for hydraulic systems is the flow restriction from the hydraulic supply unit, that limits the achievable TCP velocity, especially when driving multiple actuators simultaneously. To address this constraint, a flow-bounded control strategy is utilized. This approach is presented in detail in Lampinen et al. (2020). The selected approach is inspired by torque-bounded trajectories presented in Dahl and Nielsen (1990) and Dahl (1994), and is similar to an online method proposed recently to limit velocity in manual coordinated control (Wanner & Sawodny, 2019).

The main function of the algorithm is to dynamically scale trajectories to a velocity that is attainable for the manipulator's configuration. Due to the nonlinear nature of hydraulic systems, the attainable velocity can vary significantly depending on the manipulator configuration. To ensure that the manipulator can reach the desired velocity of the trajectory generator, the required volumetric flow rate for the hydraulic actuators must not exceed the flow rate generated by the hydraulic supply unit.

Let $\mathbf{J}_x \in \mathbb{R}^{4\times4}$ be an actuator space mapping matrix that translates joint velocities into actuator space as

$$\begin{bmatrix} \dot{q}_{\text{motor}} \\ \dot{x}_{\text{lift}} \\ \dot{x}_{\text{tilt}} \\ \dot{x}_{\text{breaker}} \end{bmatrix} = \mathbf{J}_x\dot{\mathbf{q}}. \tag{7}$$

In the case of the hydraulic motor, the velocity is simply the angular velocity of the base of the manipulator divided by the gear ratio of the ring gear and the planetary gear.

The required flow rate of each cylinder can be obtained by using

$$Q_{cylinder} = \begin{cases} A_A\dot{x}_i, & \text{when} \quad \dot{x}_i \geq 0 \\ -A_B\dot{x}_i, & \text{when} \quad \dot{x}_i < 0, \end{cases} \tag{8}$$

where $\dot{x}_i \in \{\dot{x}_{\text{lift}}, \dot{x}_{\text{tilt}}, \dot{x}_{\text{breaker}}\}$, A_A and A_B are the areas on the A- and B-sides of the hydraulic cylinder, respectively, and for the hydraulic motor by using

$$Q_{\text{motor}} = \frac{|\dot{q}_{\text{motor}}|D_m}{2\pi\eta_{vol}}, \tag{9}$$

where D_m is the volumetric displacement of the motor, and η_{vol} is the volumetric efficiency of the motor. Summing the required flow of each actuator yields the total required flow from the supply, Q_r. The scaling factor is then obtained using

$$\dot{s} = \min\left(1, \frac{Q_p}{Q_r}\right), \tag{10}$$

where Q_p is the maximum flow from the supply pump.

The algorithm is employed by the control system via a connection to the trajectory generator. In equation (2), the trajectory is a function of time. However, if we define $\dot{t} = \dot{s}$, where $\dot{s}$ is the trajectory scaling factor, we can make (2) a function of scaled time that effectively limits the trajectory to an attainable velocity. This connection is visible in Figure 11.

3.3.5 | Motion control

The motion control system used in the experiments relies heavily on learned velocity feed-forward mapping complemented by a proportional controller. The manipulator is equipped with Danfoss PVG-120 mobile proportional control valves with a significant dead-zone (approximately 30% per direction), thus making dead-zone inversion obligatory in the control design (Bak & Hansen, 2012). Moreover, it significantly improves control accuracy. For more accurate control of the manipulator, stability guaranteed model-based control methods have been shown to achieve state-of-the-art performance (Mattila et al., 2017). In Lampinen et al. (2019), such a model-based controller was proposed. Its use was demonstrated on the last link of the manipulator with a novel method of handling the nonlinearities of the pressure-compensated valves with dead zones.

In this study, velocity feed-forward learning for each valve-actuator pair was performed using the algorithm proposed in Nurmi and Mattila (2017). The algorithm identifies a feed-forward model of the valve-actuator pair by driving the actuator in a sinusoidal trajectory, while at the same time using adaptive control methods to map valve control and actuator velocity. The feed-forward model is identified in 24 distinct segments of the whole control region to accurately represent the valve characteristics.

3.3.6 | Control system verification

To demonstrate the control system's performance with dynamic trajectory tracking, a 3-DOF test trajectory was designed. This trajectory is shown in Figure 13. It consisted of five piecewise smooth segments of quintic paths generated using Equation (2), with the design time of each segment set to 8 s. However, due to the scaling of the trajectory, the timing was not absolute. The total time required to complete the trajectory was 42.4 s. The Cartesian tracking error during the trajectory was shown in Figure 14. The maximum tracking error during the trajectory was approximately 58 mm, while the mean error was 17.8 mm. Individual trajectories of each joint are shown in Figure 15, which highlights that each joint can track their respective trajectories with high precision.

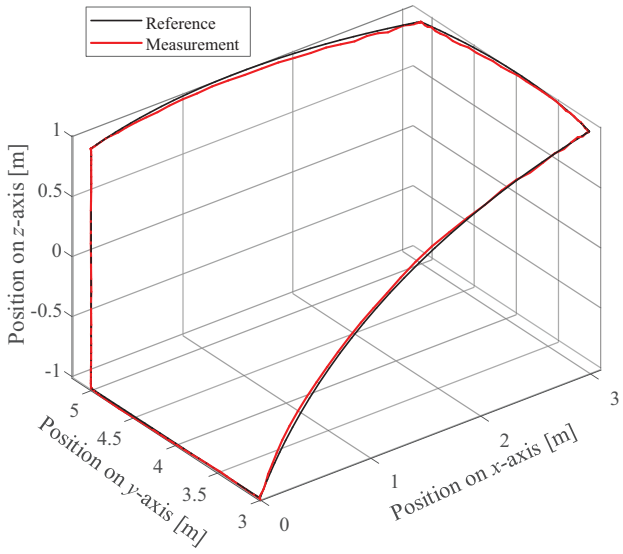

FIGURE 13 Cartesian trajectory used for control system verification [Color figure can be viewed at wileyonlinelibrary.com]

3.3.7 | Short discussion on implementing force control with force estimation

This section continues the discussion of Remark 3 on the topic of force estimation and force control. Force control of hydraulic series manipulators is not a novel concept, but due to the highly nonlinear dynamic behavior of hydraulic systems, it has remained mainly a curiosity in industrial applications, and the documented implementations limit to technical demonstrations in laboratory environment (Mattila et al., 2017). Contact identification and classification methods on the other hand have been well surveyed in (Haddadin et al., 2017).

Within the scope of this study, our aim was to create a system that requires minimal modifications to the original system and thus has less possible points of failure. With the aid of pressure sensors, similar model based control approach as proposed by Lampinen et al. (2019) could be extended to the whole manipulator. Force estimation could then be implemented using the measurable cylinder piston forces and estimated dynamics of the manipulator as proposed by Koivumäki and Mattila (2015). The more advanced model based control method could be utilized with impedance control scheme as shown in Koivumäki and Mattila (2017) to achieve the compliant and force aware contact control for a stable rock contact. A different route of utilizing force estimation could be to leave the control system untouched and use the force estimation only for contact detection and classification as well as external event detection, for example, tool slipping, rock slipping, or detection of a break instance.

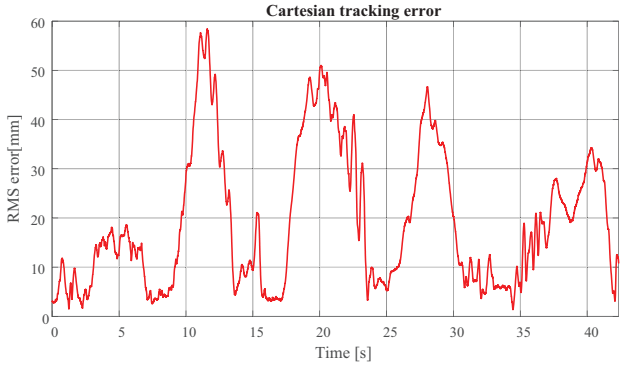

FIGURE 14 Cartesian tracking error during the experiment [Color figure can be viewed at wileyonlinelibrary.com]

4 | MANIPULATOR AND CAMERA CALIBRATION

4.1 | Manipulator calibration

An accurate forward kinematic model of the manipulator is a prerequisite for vision-based operations using absolute coordinates. Therefore, before anything else, the manipulator's internal link coordinate system, from its base to the TCP, must be calibrated using accurate external measurements, to compensate for errors in nominal link lengths and uncalibrated encoder offsets. Alternatively, the uncertainty related to the kinematic parameters could be mitigated by using eye-in-hand tracking of the TCP and relative positions (i.e., the vision system gives rock positions relative to the perceived TCP location). However, such an application could prove to be very harsh for the camera, due to the high-frequency vibrations of the impact hammer. Therefore, we opted for the kinematic calibration process instead.

All four joint axes of the manipulator are equipped with 14-bit SIKO WV58MR absolute rotary encoders, with an angular resolution of 0.022°. The external measurement device used was a SOKKIA NET05 total station laser theodolite, which provides 3D position data with sub-millimeter accuracy. A spherically mounted retroreflector was attached to the hammer tip, and its laser-indicated position together with joint encoder readings were recorded in 28 joint configurations when the boom was static.

Figure 5 illustrates the coordinate frame assignment for the boom, which was done following the Denavit–Hartenberg (DH) convention. First, the homogenous transformation from the theodolite measurement frame to the mechanical base frame of the manipulator (which is found at the intersection of its first two joints) was estimated with a circle fitting procedure (Bernard & Albright, 1994). Then, using the nominal dimensions of the boom as an initial guess, an estimate of the actual DH parameters was determined by applying the Levenberg–Marquardt algorithm to iteratively find a numerical solution that best described the boom geometry.

The resulting position residuals between the calibrated forward kinematic model and the values indicated by the external measurement device are visualized in Figure 16. The top of the figure visualizes *X*, *Y*, *Z*, and Cartesian position residuals from each individual measurement, while the bottom presents the distributions of these respective errors as a histogram. The kinematic calibration resulted in a spatial mean error of less than 10 mm and maximum errors of less

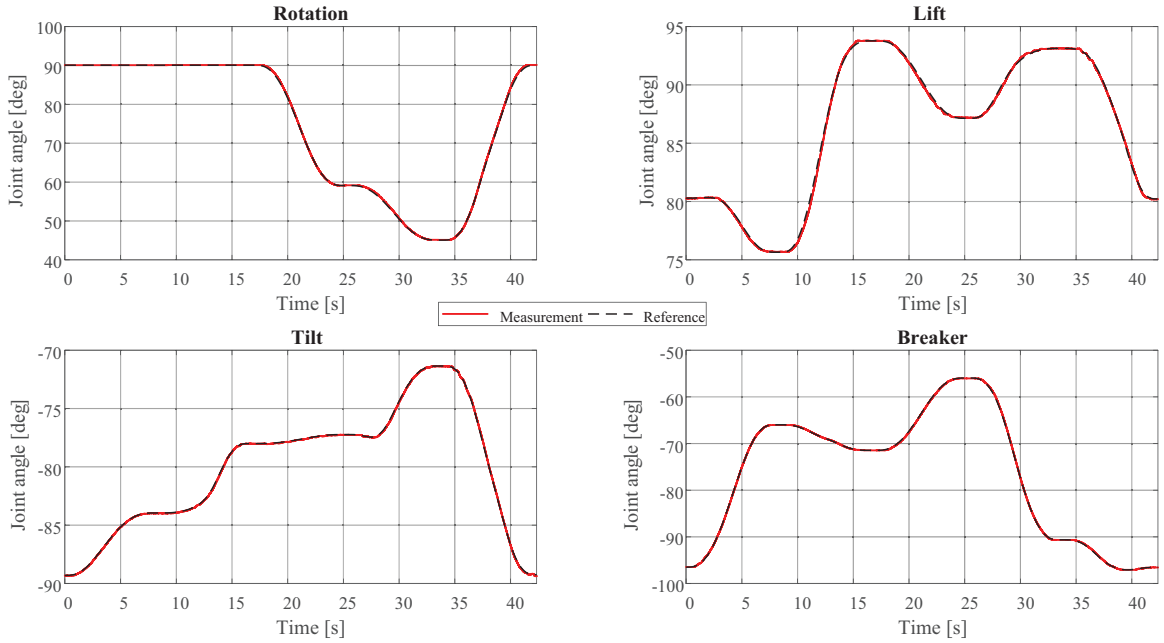

FIGURE 15 Individual joint tracking during control system verification experiment [Color figure can be viewed at wileyonlinelibrary.com]

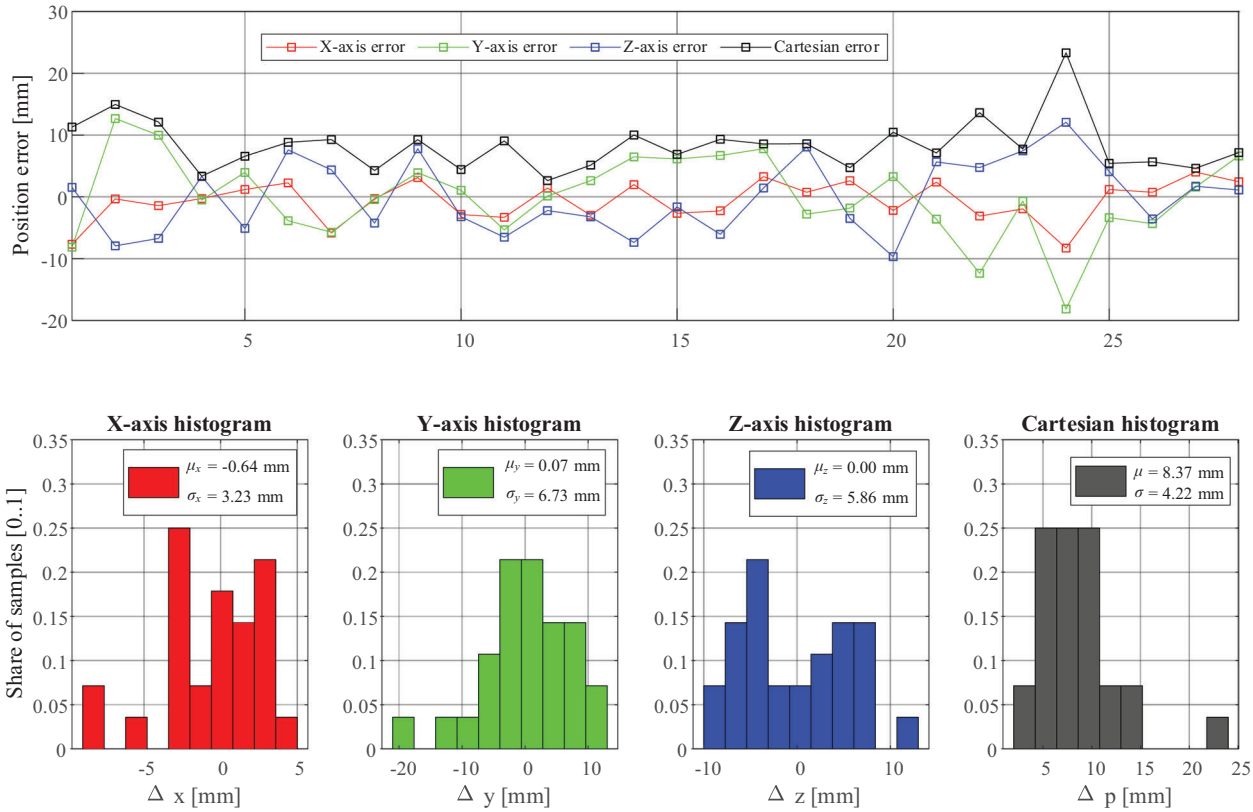

FIGURE 16 Position residuals after kinematic calibration [Color figure can be viewed at wileyonlinelibrary.com]

than 25 mm. By comparison, the diameter of the blunt tool that comes into contact with rocks is 135 mm. For a 9-ton manipulator with a reach of 7 m, this degree of accuracy can be considered impressive, and higher accuracy is likely impossible due to structural flexibilities.

As a remark, the accuracy reported here was achieved with a less than 4 year old breaker boom that has seen only light work-cycles (acting mainly as motion control platform without significant amounts of rock breaking activity) and can thus be considered

relatively intact. In an actual breaker boom plant subject to continuous stress and impacts, the repeatability of the manipulator might deteriorate over time due to wear. Consequently, the achievable absolute accuracy of the manipulator might decrease during its lifespan, despite including regular recalibration as a part of maintenance operations.

4.2 | Camera calibration

Stereo cameras have been extensively used in real-time robotic vision applications, such as detecting and measuring objects, and estimating objects' poses in a scene. The accuracy of such stereoscopic visual system relies entirely on calibration, which determines the overall performance of the system.

A stereo camera is typically composed of a pair of pinhole cameras. The camera calibration process estimates the geometric properties of the camera, as well as its pose in robot coordinate system (Forsyth & Ponce, 2002). The camera calibration parameters include intrinsic and extrinsic parameters and distortion coefficients, as illustrated in Figure 17.

Given a 3D point (X, Y, Z) in the robot coordinate system, its corresponding point (x, y, z) in the camera coordinate system and (u, v) in the 2D image plane, the extrinsic calibration parameters follow a rigid transformation between camera and robot coordinates:

$$\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} R_{3\times3} & t_{3\times1} \\ 0_{1\times3} & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}. \tag{11}$$

The intrinsic calibration parameters represent the projective transformation between the 2D image coordinates and 3D camera coordinates:

$$z \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}, \tag{12}$$

where z is the depth at the image coordinate (u, v). Combining the above two equations, a general perspective transformation can be written as

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{z} \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R \mid t \end{bmatrix}_{3\times4} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \tag{13}$$

where (f_x, f_y) is the focal length in pixels, (c_x, c_y) is the optical center in pixels, $R \in \mathbb{R}^{3\times3}$ is a rotation matrix, and $t \in \mathbb{R}^3$ is a translation vector.

4.2.1 | Intrinsic calibration

Real lenses always exhibit some radial distortion and slight tangential distortion. Camera calibration helps correct radial lens distortions and measurement errors. The Stereolabs ZED stereo camera in our system is an integrated parallel stereoscopic camera with a known baseline. Even though it comes with factory calibration, its accuracy can still be improved with recalibration. The calibration process follows a multi-plane calibration approach (Zhang, 2000), which only requires a planar pattern. Without knowing positions and orientations, calibration was performed by moving the camera with respect to the planar calibration pattern on a 27-inch 2560×1440 display. The ZED camera calibration parameters for a resolution of 1280×720 pixcel were recorded in Table 2, which contains the intrinsic parameters, focal lengths f_x, f_y, principal points c_x, c_y, and the lens distortion of both the left and right eye of the camera $[k_1, k_2, p_1, p_2, k_3]$, as well as the extrinsic parameters of the right eye with respect to the left eye of the camera, R_0 and t_0. Here, the common lens distortion can be corrected with Brown-Conrady model (Brown, 1966), which takes into account both radial distortion and tangential distortion:

$$\begin{aligned} u_d &= u_n(1 + k_1 r^2 + k_2 r^4) + \left(2p_1 u_n v_n + p_2(r^2 + 2u_n^2)\right), \\ v_d &= v_n(1 + k_1 r^2 + k_2 r^4) + \left(2p_2 u_n v_n + p_1(r^2 + 2v_n^2)\right), \end{aligned} \tag{14}$$

where $r^2 = (u_n - u_c)^2 + (v_n - v_c)^2$, u_d, v_d are coordinates in the distorted image, u_n and v_n are coordinates in the undistorted image, u_c and v_c are coordinates of the distortion center, k_1 and k_2 are radial distortion coefficients, and p_1 and p_2 are tangential distortion coefficients.

4.2.2 | Extrinsic calibration

The extrinsic parameters are determined by how the camera is positioned in the robot coordinates. A point in the camera coordinate frame depicts its position with respect to the optical center of the left eye of the camera. For the robot manipulator, it is more useful to know where this point is relative to the robot base in the robot coordinate frame, which coincides with the world coordinate frame. Both the camera and the robot coordinate frame follow the right hand rule, as shown in Figure 5.

For a stationary camera and robot base, any objects in the camera coordinates should retain the same geometry and scale after transformation into robot coordinates. This geometric transformation aligns every corresponding point of two-point cloud sets; this is referred to as rigid transformation, which is expressed by Equation (11). The approaches to find rotation R and translation t can be categorized as singular value decomposition (SVD)-based (Arun et al., 1987; Ho, 2013) and quaternion-based (Horn, 1987; Horn et al., 1988; Walker et al., 1991). The SVD-based method was adopted to obtain the highest possible level of accuracy and stability.

Let the points in camera coordinate be $C = \{c_i\}$, and $c_i = [x_i, y_i, z_i]^T$ in 3D, where $i = 1, 2, \ldots N$ and N is the number of points and corresponding

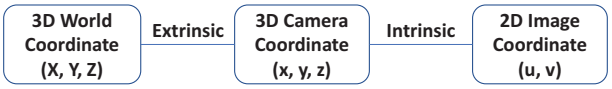

FIGURE 17 Camera calibration process [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 2 Stereo camera parameters

ZED	f_x	f_y	c_x	c_y	Distortion coefficient [k_1, k_2, p_1, p_2, k_3]
Left	700.79	700.79	634.822	356.993	[−0.176, 0.029, 0.00196, −0.00044, 0.00]
Right	700.71	700.71	626.699	356.066	[−0.172, 0.027, 0.00164, 0.00016, 0.00]
R_0	[−0.00659, 0.01328, −0.00013]				
t_0 [mm]	[−120.002, 0.00, 0.00]				
R	[−0.019, −0.003, −0.024]				
t [m]	[−0.480, −4.491, −3.209]				

points. Then let the corresponding points in the robot coordinates be $W = \{w_i\}$, and $w_i = [X_i, Y_i, Z_i]^T$ in 3D, which can be expressed as

$$\begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} = R_{3\times3}^{-1}\left(\begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} - t_{3\times1}\right). \tag{15}$$

According to the SVD approach,

$$[U, S, V] = SVD(CW^T), \tag{16}$$

where C and W are the $3 \times N$ matrices that have $c_i - \frac{1}{N}\sum_{i=1}^{N} c^i$ and $w_i - \frac{1}{N}\sum_{i=1}^{N} w^i$ as their columns, respectively, and U and V are orthonormal matrices from which we obtain

$$R = VU^T. \tag{17}$$

Subsequently, the translation vector t can be computed as

$$t = \frac{1}{N}\sum_{i=1}^{N} c^i - R\left(\frac{1}{N}\sum_{i=1}^{N} w^i\right). \tag{18}$$

Unlike intrinsic calibration, which was performed indoors, the actual extrinsic calibration was conducted outdoors at a rock breaking field. The calibration workspace in the field consisted of measurement equipment, the grizzly, markers, the overhead ZED camera, and the Rambooms manipulator. For successful extrinsic calibration, the following conditions were fulfilled:

- The intrinsic calibration of the ZED camera was completed;
- The manipulator's forward kinematics model was calibrated;
- The marker positions in camera coordinates were accurately measured;
- The corresponding marker positions in robot coordinates were accurately measured.

Extrinsic calibration was initiated by distributing markers into the workspace in such a way that flat red markers appeared around the middle of the camera's field of view. To achieve the best possible contrast between the markers and the background, lower exposure for image acquisition was set. An example of the acquired image (at 1280 × 720 resolution) is shown in Figure 18, where there are six markers in the scene. Next, the process for marker segmentation was performed to the corresponding point cloud in the camera coordinates using a color mask to filter all objects except the markers. After filtering, what remained of the scene were the markers, presented in the form of clusters of points, as shown in Figure 19. A k-means clustering algorithm was used to locate all markers' centroid positions and thus obtain their positions in the camera's coordinates.

For measuring marker positions in robot coordinates, multiple approaches were available. A quick and effective approach was to take advantage of the calibrated Rambooms manipulator to manually align the centroid position of each marker with the manipulator's TCP, so the manipulator could convert each marker position to robot coordinates. For this method, robot accuracy must be taken into account. Once all corresponding camera and robot coordinate positions were measured, measurement data was validated and represented in homogeneous coordinates. Finally, the extrinsic parameter rotation matrix R and translation vector t were calculated according to Equation (17) and (18), respectively. These numerical values are presented in Table 2, together with the camera intrinsic parameters.

4.2.3 | Verification of vision system accuracy

The accuracy of the vision system was determined with the ZED camera's inherent accuracy, intrinsic and extrinsic camera calibration methods, and measurement errors during camera calibration. The

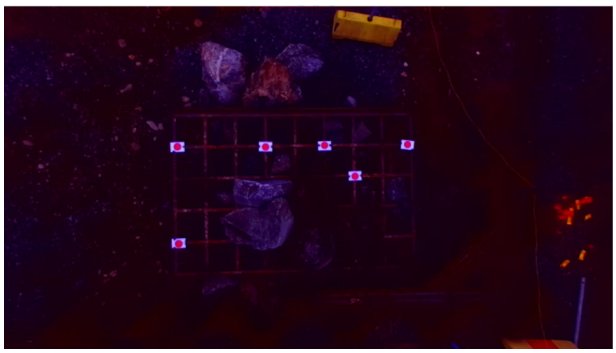

FIGURE 18 A view of red markers acquired at low exposure [Color figure can be viewed at wileyonlinelibrary.com]

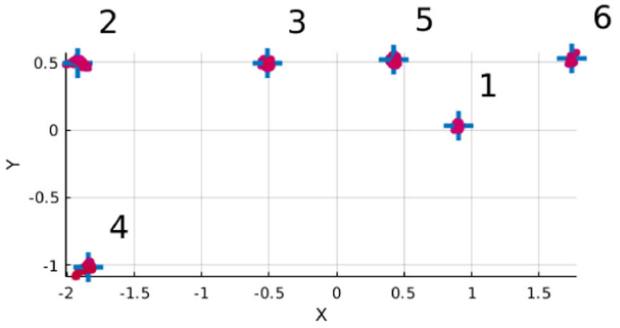

FIGURE 19 An example of red markers in a point cloud in the camera coordinates. The blue crosses denote the markers' centroid positions [Color figure can be viewed at wileyonlinelibrary.com]

ZED camera is a binocular pinhole camera with an operating range of up to 20 m, but its depth accuracy decreases when the distance between the camera and the target increases. The accuracy test was conducted with four to five meters between the grizzly and the ZED camera. The target objects were the red markers used for calibration. The calibrated intrinsic and extrinsic parameters of the ZED camera were used to obtain the estimated target object positions in robot coordinates. These estimated marker positions were then compared to corresponding data given by the manipulator. In total, 24 measurements were performed. Figure 20 illustrates position residuals, where the maximum deviations are 18.4, 22.4, 67.5, and 67.19 mm along *X*, *Y*, *Z* axes and Cartesian, respectively. SDs in the *X*, *Y*, *Z* axes and Cartesian are 10.0, 10.5, 27.9, and 15.14 mm, respectively. Considering the 135 mm diameter of the manipulator's blunt tool, as well as the grizzly's grid size of 400 × 400 mm, the position errors are acceptable.

5 | EXPERIMENTS AND RESULTS

5.1 | Rock detection

Rock detection and localization is a crucial step in autonomous breaking. During the data collection phase, we collected a total of 4733 images of the breaking scene with the grizzly and rocks visible. Ninty percent of them were used for training and validation data, while the remaining 10% were used for model and system-level testing, which ensured the final experiments could be carried out seamlessly.

The maximum amount of rocks the grizzly can hold depends on the size of rocks. During the final rock breaking experiments, the amount of rocks in each experiment varied from 6 to 12. An example of a scene with 12 rocks is presented in Figure 21, where the detection speed was 85 ms. Once a rock was detected, it was numbered and enclosed in a purple bounding box. The total sum of detected rocks is shown in the upper left corner of the image.

To further enhance the robustness of the rock detection process, we set a region of interest (ROI) indicated with white rectangles in the image coordinate frame. Rocks detected outside the ROI were ignored. Whether a rock lay inside or outside the ROI was determined by the centroid position of its bounding box. In addition, we ignored rocks smaller than the grid openings of the grizzly, as well as rocks laying beneath the grizzly's upper surface.

5.2 | Position for rock breaking

The positions and orientations for rock breaking presented in Figure 21 were estimated based on the reconstructed 3D surface point cloud of each rock. As the software architecture in Figure 6 depicts, this was a long process that began with 3D reconstruction of the environment using the detected 2D images. Within each region, every pixel was reprojected onto a corresponding point in the camera coordinates with position values (*X*, *Y*, *Z*) and RGB color codes. We could thus obtain detected rocks in the form of 3D regions in a dense point cloud. This process is detailed in Section 3.2.3. Next, the reconstructed 3D regions were transformed from camera coordinates to robot coordinates. This result is shown in Figure 22. Finally, the process for estimating the position and orientation required for rock breaking could be initiated. A more detailed description of each of these processes is provided in Section 3.2.4 and 3.2.5, respectively.

To validate the results, the breaking positions for the aforementioned 12 rocks are visualized in Figure 23, which verifies the correctness and effectiveness of the position values online in Figure 21. The positions for guiding the manipulator's blunt tool are marked as red spots on the surface of each rock (including partially occluded ones). A 3D viewer was implemented for live monitoring purposes using the Point Cloud Library (PCL) in C++.

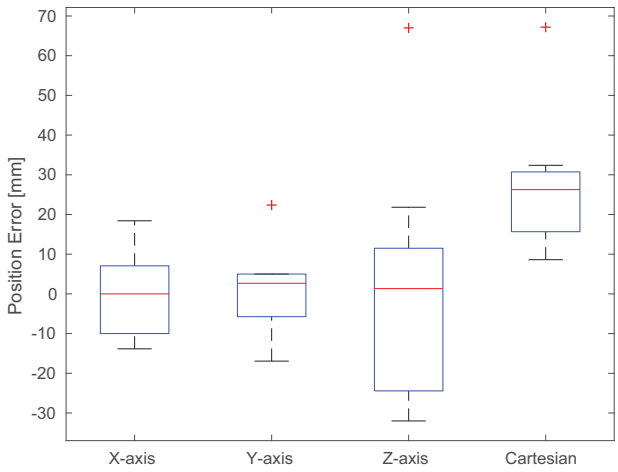

FIGURE 20 Box plot of position residuals in the *X*, *Y*, *Z* axes and Cartesian. The boxes show the 25th and 75th percentiles of residuals, while red lines represent medians, whiskers represent the minimum and maximum values, and the red "+" symbols signify large residuals regarded as outliers [Color figure can be viewed at wileyonlinelibrary.com]

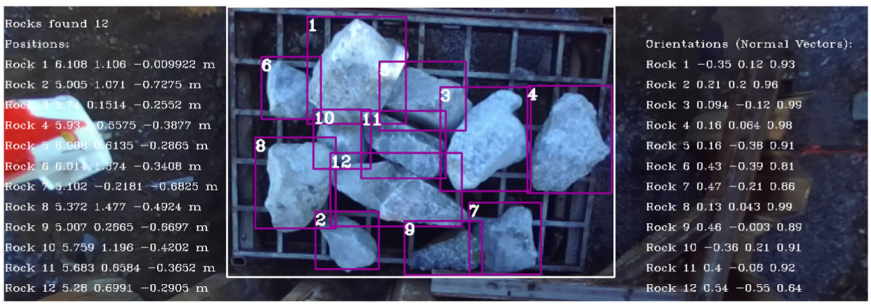

FIGURE 21 Real-time view of the rock detection scene. The positions and orientations for rock breaking are aligned to the left and right side accordingly [Color figure can be viewed at wileyonlinelibrary.com]

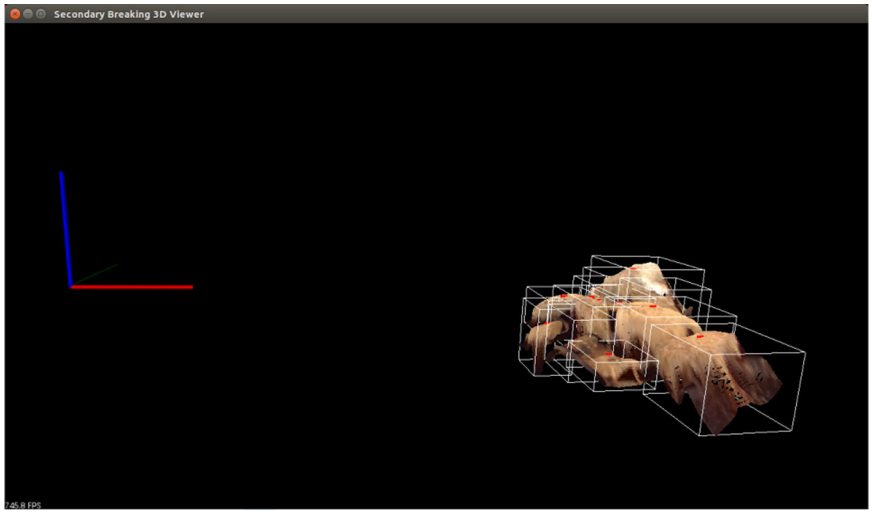

FIGURE 22 Real-time view of the detected rocks in a 3D point cloud in robot coordinates (the *X*, *Y*, and *Z* axes are marked in red, green, and blue, respectively) [Color figure can be viewed at wileyonlinelibrary.com]

5.3 | Description of the autonomous operation evaluation

The experiments reported in this manuscript serve as a technical demo. The experiments herein represent real use cases of the manipulator and demonstrate the potential for increased automation of such systems. The experiments consisted of autonomous rock breaking ranging from 1 to 10 min long, during which there was no human intervention with the autonomous operation. Before each experiment, 5–10 rocks of various sizes and shapes were laid on the grizzly. During the experiment each break attempt were recorded. After the experiment, success rates and break pace were calculated.

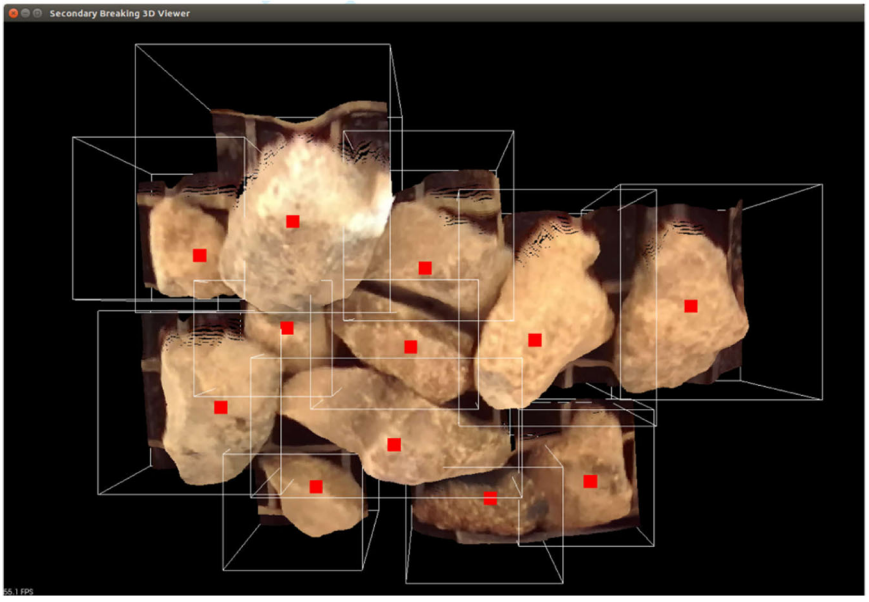

FIGURE 23 Real-time view of the detected rocks with breaking positions indicated in red dots [Color figure can be viewed at wileyonlinelibrary.com]

The break pace describes how many break attempts were performed within a 1-min span.

Each of the experiments were performed without any human intervention during the process. The system was equipped with additional safety features, such as a dead man's switch for the breaker operation and an emergency stop for the entire manipulator to ensure safety during the experiments. Fortunately, there was no need for either of these features during the experiments. The safety features included in the control system already prevented direct hits against the grizzly, and the manipulator was operated in a limited area above the grizzly. The logic of the autonomous operations during the experiments follows what is described in Section 3.3.1 and in Figure 12. The process was repeated until no rocks remained on the grizzly or the operation was halted by the operator.

Remark 4. The rocks used in the experiment were granite from a nearby construction site. This is not a typical material in this application, which should be considered when analyzing the results. Granite can have compressive strength values of over 200 MPa, which makes it very difficult to break. This might lead to lower overall success rates than what could be achievable in an actual environment with more brittle material.

5.4 | Results from the autonomous breaking experiment

The autonomous breaking experiment consisted of approximately 47 min of autonomous operation with varying numbers of differently sized rocks laid on the grizzly. The breaking was considered successful if the rock broke into two or more pieces that were at least 25% of the original volume of the rock being broken, or if the hammer pushed the rock through the grizzly. Break success rates were evaluated visually from recorded videos. The average break success rate was 34.2%, while the break pace was approximately 3.30 attempts per minute. The experiments were conducted during a very bright weather and objects in the scene suffered from over exposure and shadows which notably differed from the conditions in our data set, which resulted in degraded performance of the VPS. The average Recall of the VPS during the experiments was approximately 75%. Results from each individual experiment are gathered in Table 3. All the experiments presented here were conducted without using the rock surface normals as breaking guides; instead, the hammer was kept at a 90-degree angle with respect to the ground at all times.

During Experiment 6, we implemented a change to the autonomous operation so that after every third attempt, the manipulator moved aside to allow a better view of the grizzly for the stereo camera. Despite the addition of this extra step, the break pace remained almost identical as the manipulator's movements were also sped up. In Experiment 6, a sharp angled rock was very close to the area where the manipulator was moved after every third attempt; consequently the system attempted to break that specific rock always first. Due to the difficulty caused by the shape of the rock, the success rate of Experiment 6 was lower compared to other experiments.

As seen in Table 3, the success rate between experiments varied considerably. This was caused mainly by the fact that the rocks on the grizzly in each experiment were dissimilar in shape and size. Sharp edges were identified as causing the most problems for the system, which is owed to the relatively low stiffness of the manipulator. An especially problematic feature of the manipulator was the backlash of the rotation joint. The rotation of the manipulator is controlled with a hydraulic motor attached to a planetary gear and a ring gear, which together contribute as a significant source of backlash. Because of this, the manipulator slipped on inclined surfaces easily.

The reference and measured positions of the manipulator's TCP during Experiment 8 are visualized in Figure 24, where the red lines indicate the measured trajectory while black dashed lines represent the reference trajectory. Accurate path following can be observed in this figure. A closer look at the first 100 s of the experiment at the

TABLE 3 Results from autonomous breaking experiments

	Duration	Break attempts	Successful breaks	Success rate (%)	Break pace (attempts/min)
Experiment 1	3 min 50 s	13	7	53.8	3.39
Experiment 2	4 min 15 s	14	5	35.7	3.29
Experiment 3	7 min 30 s	26	8	30.8	3.47
Experiment 4	1 min 10 s	4	1	25.0	3.43
Experiment 5	4 min	12	5	41.7	3.39
Experiment 6	5 min 30 s	19	2	10.5	3.84
Experiment 7	5 min 30 s	16	7	43.8	3.15
Experiment 8	5 min	18	5	27.8	3.70
Experiment 9	10 min	33	13	39.4	3.42
Total	47 min	155	53	34.2	3.30

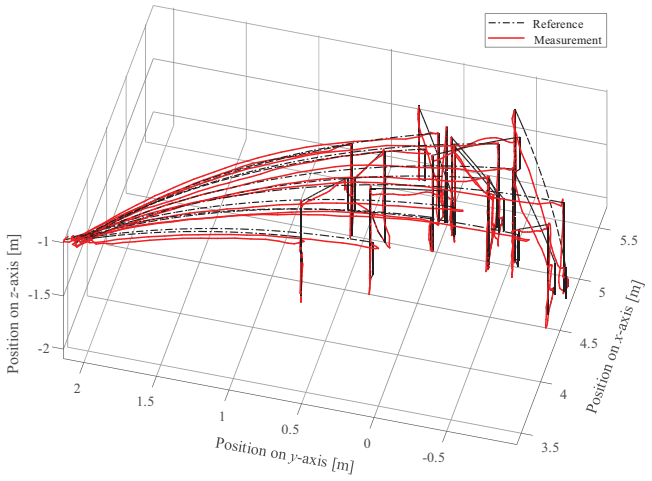

FIGURE 24 Trajectory during 5 min of autonomous operation in Experiment 8. Continuous visits at the left corner represent the manipulator moving aside to allow better view for the stereo camera over the grizzly [Color figure can be viewed at wileyonlinelibrary.com]

individual joint tracking level is shown in Figure 25 while RMS tracking errors over the same time frame are shown in Figure 26. The figures indicate the accurate tracking of each joint and demonstrate that the average Cartesian RMS tracking error over the whole experiment was only 57.9 mm. The RMS error was the largest when the controller detected broken rock and began to raise the TCP, preventing it from hitting the grizzly due to the inertia of the manipulator. The interaction with the rocks before the break attempt also caused tracking error due to the trajectory being set below the rock's surface to achieve pressure against the rock.

A close-up of an individual break attempt is shown in Figure 27 in which TCP height and lift valve control are shown in parallel to give better understanding of the actual breaking process. In the beginning, the manipulator moved to the rock. Then, control was continued in an open-loop manner, mimicking manual operation. Pressure against the rock was first built up using the dual cylinders of the lift joint of the manipulator. After pressure against the rock was ensured, the hydraulic hammer was activated. At this point, the manipulator began slowly moving downward until the rock was broken, which can be observed as a sudden loss of opposing force from the rock. In turn, this caused the manipulator's quick acceleration. When the manipulator descended below a threshold height, the rock was considered broken and the manipulator was raised back up. Due to the large inertia of the 2700 kg hydraulic hammer, the manipulator kept descending below the threshold height despite the fact that the control valve had been changed to the opposite opening. Therefore, for safety reasons, the threshold height must be set higher than the actual grizzly height to avoid impact with it. In our experiments, this height was set 50 mm above the grizzly. Due to the large inertia of the manipulator, the TCP could occasionally hit the grizzly after a rock was broken, but the hammer operation was halted automatically in advance so that no damage to the grizzly or the boom could happen.

In this study, we limited our focus on the use of only joint angle encoders for the sake of applicability to industrial applications with minimal need of retrofitting the system. However, if we consider the break attempt shown in Figure 27, it is evident that the use of pressure sensors would prove beneficial in detecting the instance a rock is broken; during this time, the pressure inside the lift cylinders collapse due to the loss of opposing force from the rock, thus making the break detectable via pressure sensors. The impacts from the

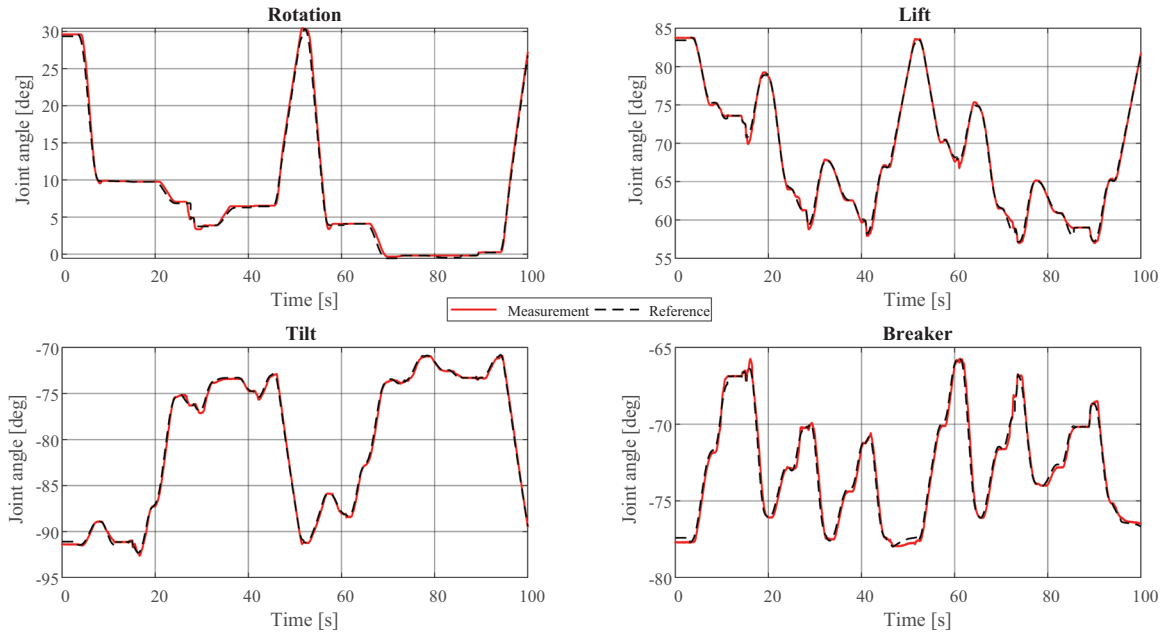

FIGURE 25 A snippet of individual joint angle tracking during break Experiment 8 [Color figure can be viewed at wileyonlinelibrary.com]

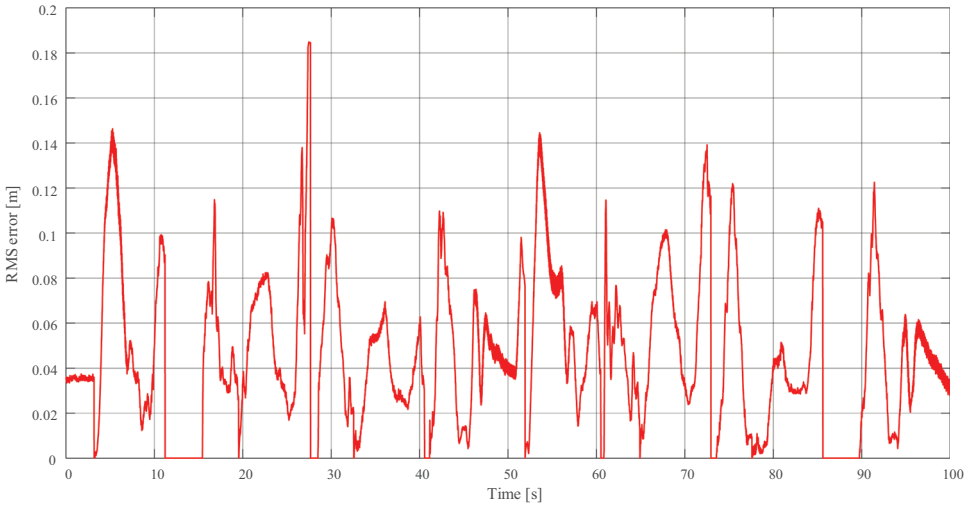

FIGURE 26 Cartesian tracking errors during Experiment 8. The mean error over the whole experiment was 57.9 mm, while the maximum error was 202 mm. The maximum error occurred during a breaking process, immediately after a rock was broken and the controller was attempting to stop the movement of the TCP toward the grizzly due to the sudden loss of opposing force from the rock. RMS, root mean square [Color figure can be viewed at wileyonlinelibrary.com]

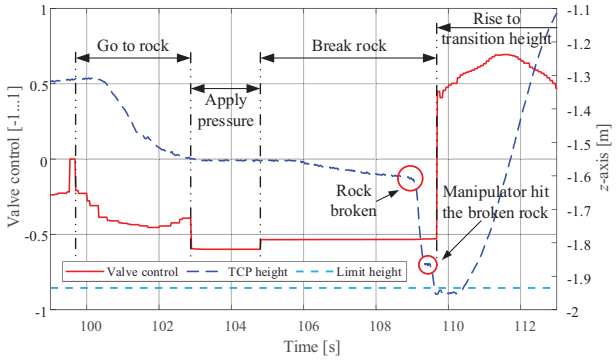

FIGURE 27 Individual break attempt from Experiment 9. The left *y*-axis shows the valve control of the lift valve used to maintain pressure against rock during the breaking process. The right *y*-axis (labeled *Z*-axis) shows the TCP's height. When the breaking begins, valve control is lowered slightly to limit the jerk of the manipulator upon the sudden loss of opposing force when the rock is broken. After approximately 0.5 s after the rock was broken the manipulator hit one half of the broken rock that is seen as a pause in the descent of the TCP in the plot. TCP, tool center point [Color figure can be viewed at wileyonlinelibrary.com]

hydraulic hammer can also be detected from pressure spikes. Using such data, the breaking process can be controlled more precisely.

6 | SUGGESTED IMPROVEMENTS

The performed experiments in our simplified field test environment served as a feasibility study providing valuable insight on the automation of the secondary breaking tasks in grizzly applications. However, the setup also limited the number of experiments practically possible to perform, due to a large number of auxiliary tasks required for each experiment. Yet, the simplified field test environment served us well and helped at identifying several key challenges and shortcomings of the proposed system. Further improvement of the proposed system would benefit from testing environment with continuous ore flow which would mean preferably an operational mine. However, these are continuously running highly optimized production facilities, where production downtime can lead to large costs. Arranging an experimental setup in such conditions without affecting the mine operation is a challenging task that requires a lot of planning, coordination, and development to achieve a mature enough research platform that can be used in cooperation with the mine operation. Such environment would prove fruitful for gathering data for performance analysis between an autonomous system and a human operator. Similar study comparing forwarder operators with boom tip control versus traditional control were conducted in Manner et al. (2019).

This section presents improvements to the proposed system based on the insight gained from the experiments. The major improvements to the rock breaking system are aimed to improve the robustness and effectiveness of the system under more complex and realistic environments. The herein proposed changes enhance the tactical layer of the rock breaking system, bringing it one-step-closer to practical implementations. These improvements are seen as way to overcome the following shortcomings of the proposed system identified during the experiments: (1) the break position selection was able to select a break position on inclined surfaces under specific conditions. (2) the system was unable to distinguish the difficulty of a break attempt beforehand in any way. (3) reorientation of difficult rocks was not considered as a strategy for more difficult rocks. (4) detection of the break instance relied only on position measurements, which yielded slow reaction to the dynamically fast break instance.

6.1 | Break point selection

A few flaws in the proposed method for break position selection were identified in the experiments, that should be addressed in future research. First, the constraint that the break position must locate within a rectangular region quarter the size of the bounding box enclosing the boulder does not always yield the best break position. For instance, if the boulder has its centroid of mass near one of its edges and is shaped like an off center pyramid, the highest point within the search area would yield a break position on a slanted surface. Based on the experiments, those conditions result with a high likelihood in an unsuccessful break attempt. Second, due to the physical constraints of the manipulator, aligning of the rock hammer is not possible in most cases. Therefore, finding the surface normal based on the break point may not be the best approach. Instead, an alternative method could be investigated in the future. Rather than selecting the angle for impact based on the break point, the break point should be selected based on the surface properties of the rock to minimize the possibility of slipping and also avoid having to align the manipulator for each rock separately.

To fulfill these conditions, suitable break locations may be obtained by first calculating normal vectors for each point of a point cloud representing a single rock. Then, the point cloud is filtered based on the normal vectors, to remove points belonging to inclined surfaces. After filtering, the point cloud is left with points belonging to flat surfaces. Then, a suitable break position can be obtained by finding a point with most points in its local neighborhood, i.e., within a specific radius around the point. Figure 28a,b illustrate the results of this method. The radius is here set to 65 mm, which is the same as the radius of the blunt tool of the hydraulic hammer.

6.2 | Per rock difficulty estimation

The inability to make any difference between rocks on the grizzly was identified as an adverse property of the evaluated VPS. Without any information regarding the difficulty of break attempts, the autonomous system was selecting it's targets only based on the Euclidean distance metric. However, this was identified as overly simplified approach, as the strategy for difficult rocks often involves manipulating the rocks into different poses.

The above described method for the break location identification can be leveraged for estimating the difficulty of a break attempt at the same time. By analyzing the surface properties of the local neighborhood of the break location its flatness can be estimated and used as a quality metric. Break points with more points in its local neighborhood in the filtered cloud indicate flatter surfaces, while low number of points in the local neighborhood indicate sharp-edged surface, as point belonging to inclined surfaces are removed. As observed in the experiments sharp-edged and inclined surfaces are much harder to break.

Other criteria related to the per rock difficulty include, for example, the height of the rock in relation to its cross-sectional area and the horizontal distance of the break point and the centroid of the rock. Other criteria may exist as well, such as the overlapping percentage, but the fine-tuning and finding correct relations between each criteria becomes more complex with more criteria. The main goal of the difficulty estimation is after all determining whether a rock may be broken as such or it should be manipulated into a different pose first. Figure 28c has the difficulties estimated for each break location based on the three first-mentioned criteria. Difficulty scores below 10 indicate a typical difficulty, while scores above 10 indicate a challenging rock, that may need to be manipulated into a different pose before a break attempt. Defining a meaningful metrics for the difficulty estimation requires further testing to obtain a balanced relation between different difficulty criteria, and to validate the results.

6.3 | Break instance detection

Another shortcoming of the experimental setup was the naive approach for detecting break instances, or more specifically, the lack of such system. Therefore, we acknowledge the necessity to describe a few approaches that could be leveraged to detect the break instances. To detect the shattering of the target rock, an intuitive

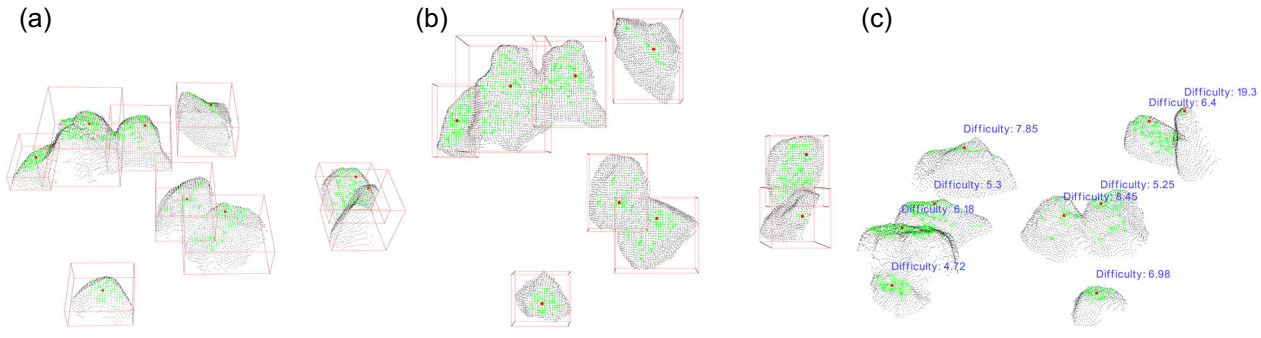

FIGURE 28 Visualization of the results of the second break location selection method. The suitable areas with normal vector pointing up are drawn with green, while the selected break points are marked with red points. (a) Break locations from side, (b) break locations from above, and (c) break difficulty estimation [Color figure can be viewed at wileyonlinelibrary.com]

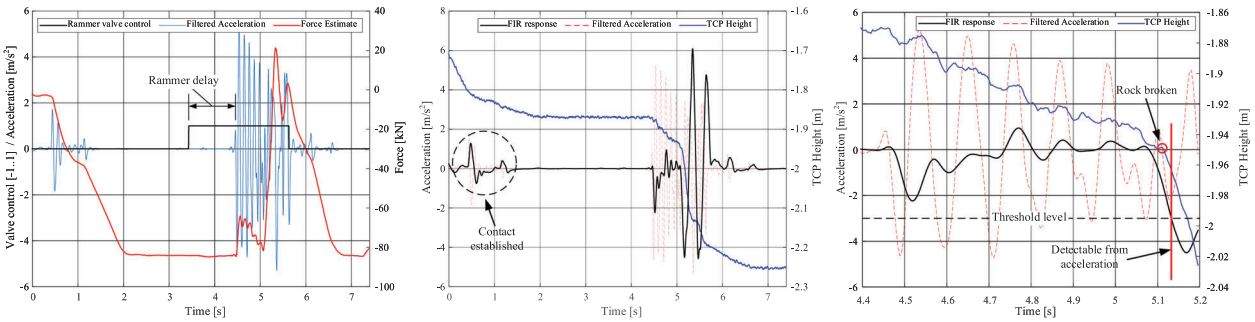

FIGURE 29 Demonstration of the break instance detection using external force estimation and acceleration measurements [Color figure can be viewed at wileyonlinelibrary.com]

method is to observe the forces of the manipulator and detect a sudden loss of external opposing force at the TCP. However direct contact force measurement is not practical due to the harsh application, that is typically of of question for fragile force/torque sensors. Instead, indirect approaches are required to estimate the external force. Takahashi and Monden (1999) proposed strain gauges at the chisel of the hammer to detect the external force and ultimately the break instance. However, the chisel is subject to such harsh use that the longevity of the strain gauges is susceptible at best. In Section 3.3.7, force estimation based on model based dynamics compensation and pressure signals was discussed. Such approach is an effective way to obtain a force estimate for the break instance detection, but it requires sophisticated modeling of the dynamics of the manipulator as well as multiple pressure sensors (in this case six), both typically avoided in commercial applications. A simpler approach could be to monitor the pressure of only one side of one cylinder and detect rapid changes to identify the break instance. Alternatively, accelerometer could be utilized for the same purpose.

Figure 29 visualizes measurements from a successful break attempt performed manually. The break instance is here detectable by two methods. First, we have implemented the force estimation within the manipulator controls, and second, we utilize acceleration measurements gathered using a Novatron G2 IMU. The acceleration in the direction of the hammer is first filtered using a high-pass filter to remove the effect of gravity. Then, a FIR filter with a window length of 120 ms is used to detect the break instance. The frequency of the hydraulic hammer is roughly 8 Hz, which yields a near zero response from the FIR filter at the hammer operating frequency. However, when the rock is broken, the response will show a large negative value, due to one half of the acceleration spike being lost. This method enables us to detect a successful break within approximately 60–90 ms.

6.4 | Improved autonomous operation pipeline

The suggested improvements lead to an improved operational pipeline for the autonomous system that has tactical tools to handle more versatile and realistic situations and environment. The improved break point selection is foreseen to improve the success rate of breaking as the break points are selected more carefully on locations that inherently minimize the possibility of manipulator slipping on the rock surface, or the rock moving away. Moreover, the ability to distinguish challenging rocks from the grizzly further enhances the success rate as the system can first focus on easier rocks and then attempt the harder cases after it has attempted to reorient

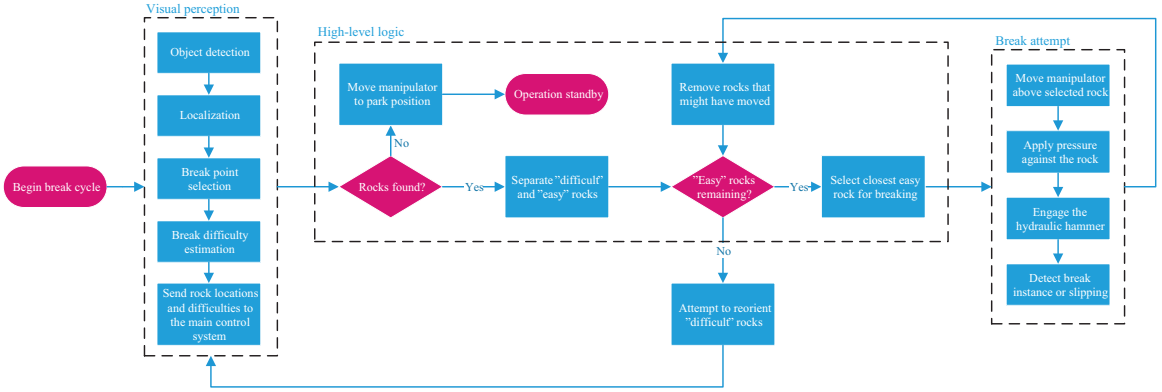

FIGURE 30 An improved autonomous operation pipeline based on the tactical layer improvements [Color figure can be viewed at wileyonlinelibrary.com]

them first. To manipulate difficult rocks into better poses, we propose that the manipulator is driven through a trajectory that goes trough the centroid of each difficult rock. Other manipulation tactics can be applied here as well, for example, the rock may be moved toward the largest unoccupied area near it, or as a last resort just to the side of the grizzly for later manual treatment by a human operator.

The break instance detection can be seen as a feature that increases safety of the operation and at the same time reduces time between break attempts. The reduced time between the detection of a break instance is very critical for stopping the manipulator's downward movement after a sudden loss of opposing force from the rock being broken. Similar methodology can be applied to detect other relevant information about the break attempt, for example, number of impacts from the hydraulic hammer, or the slipping of the manipulator.

Figure 30 illustrates a simplified operational pipeline possible with the improvements suggested in this section. The pipeline is simplified in a sense that it only covers the operations in a high level leaving smaller details out to just give clear visualization.

7 | CONCLUSIONS AND FUTURE WORK

In this paper, we presented a novel autonomous robotic rock breaking system that was verified in a full-scale real-world environment. The proposed system is built on two developments: (1) a commercially available rock breaker boom that is instrumented with high-precision joint angle encoders used for our robotic control system deployed on a dSpace MicroAutoBox 2 real-time system, and (2) a 3D VPS incorporating YOLOv3 infrastructure. Recent leaps in the performance of deep learning models enabled us to design a 3D vision system for resolving real-world object detection problems. The proposed integral system enables completely autonomous rock breaking, which serves to increase automation in the mining industry.

The main contribution of this paper is the full-scale demonstration and integration of an autonomous system for breaker boom operation in grizzly applications. Such a system has been proposed as early as the 1990s, but to the authors' knowledge, it has not been demonstrated before this study. The vision system can achieve rock detection with an average precision of 97.61%. The manipulator control system has dynamic accuracy of 60 mm in free space operation, which is sufficient for the intended application. Based on the data gathered by the vision system, the autonomous control system finds the best rock to break and can sequence the rock breaking, even in a cluttered and dynamic scene.

The proposed system was shown to be capable of autonomous operation without any need for human intervention. In the experiments, the system was able to both recognize and localize oversized rocks on the grizzly, move the manipulator in contact with the rocks, and engage the hydraulic hammer to reduce the size of oversized rocks. The shape of the rocks was identified as a crucial factor in successful breaking, especially with rocks slanted perpendicular to the rotation of the manipulator, which would slip away from the rock during the breaking process when pressure was applied against the slanted rock. The main reason for this is the backlash in the rotation joint of the manipulator. In the authors' opinions, the break success rate could be substantially increased even in manual operation if this problem is addressed and considered in the manipulator's design. Despite these challenges, our proof-of-concept system was able to achieve a success rate of 34%, which can be considered an adequate result for a system that can operate continuously without interruption at a pace of over 3 break attempts per minute. It is also worth noting that even a human operator will not achieve a 100% success rate, as the same challenges of rocks or the manipulator slipping away will persist regardless of the operator. However, a human operator can most likely react and adapt to these challenges more rapidly by attempting to break different spots on the rock based on observing to which side the rock is rolling or moving. Teaching the manipulator control system how to make such observations may be explored in future research to enhance the performance of the proposed system.

The results from the preliminary study of autonomous rock breaker operation presented in this paper are promising and highlight the system's technological readiness. The road from technical demo to commercial product is long and requires a lot more testing, but the main challenges of employing such a system can be overcome, as shown in this paper. In the experiments, no operator interference was required and no unexpected behavior was encountered. The system had a good success rate in breaking large and sturdy rocks, while smaller rocks with sharp shapes caused difficulty for the system. Nevertheless, the benefit from an autonomous rock breaker is seen in the more consistent and tireless nonstop operation that is only achievable with robotic operations. The occasional situations when the system requires human intervention for aid in the breaking is seen negligible, as operators are foreseen to be able to simultaneously monitor multiple booms.

Future work on this system should focus on further developing the individual components of the system, as well as aim for wider and more practical testing. More comprehensive experiments in a relevant environment are required to obtain quantitative information to analyze and compare the performance of the autonomous system to experienced operators. An interesting path for further research is to implement skills that mimic and exploit operating strategies that experienced operators use to achieve higher success rates, for example, manipulating difficult rocks into different poses for easier breaking. It is worth noting that in different breaker boom applications, especially with gyratory crushers, the raking and manipulation of rocks to prevent and remove blockages is almost as widely used of a strategy as the actual breaking. Different operating strategies for these applications should also be considered in future research. In an ideal scenario, the system could be deployed in a real mine as a secondary system for testing and prototyping purposes. On the subsystem level, the VPS requires more training data on different situations to cover a broader range of environmental conditions, for example, snowfall, rainfall, fog, dust, and artificial lighting. In addition,

the currently used stereo camera is unsuitable for the intended harsh outdoor conditions of the target application. As such, alternatives should be considered and designed from scratch, if necessary. Another interesting path for future research lies in the control of the breaker boom—more specifically, contact control with the rocks. To control the force applied to the rocks, the use of more sophisticated control methods must be investigated. A promising way to achieve higher precision control lies in nonlinear model-based control.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Github at https://github.com/epoc88/SecondaryBreakingDataset.

ORCID

Santeri Lampinen http://orcid.org/0000-0003-1201-9027
Longchuan Niu https://orcid.org/0000-0003-4996-5519
Jouni Mattila https://orcid.org/0000-0003-1799-4323

REFERENCES

Al Hakim, E. (2018). *3d yolo: End-to-end 3d object detection using point clouds* (Master's thesis). KTH, School of Electrical Engineering and Computer Science (EECS).

Arun, K. S., Huang, T. S., & Blostein, S. D. (1987). Least-squares fitting of two 3-d point sets. In *IEEE Transactions on pattern analysis and machine intelligence*, PAMI-9 (Vol. 5, pp. 698–700).

Bak, M. K., & Hansen, M. R. (2012). Modeling, performance testing and parameter identification of pressure compensated proportional directional control valves. In *Proceedings of the 7th FPNI PhD Symposium on Fluid Power*, June 27–30, 2012, Reggio Emilia, Italy (pp. 889–908).

Bentley, J. L. (1975). Multidimensional binary search trees used for associative searching. *Communications of the ACM*, *18*(9), 509–517.

Bernard, R., & Albright, S. (1994). *Robot calibration*. Springer.

Boeing, A. (2013). A remotely operated robotic rock breaker with collision avoidance for the mining industry. In *30th International Symposium of Automation and Robotics in Construction and Mining (ISARC 2013) Proceedings*, IAARC, (pp. 875–884).

Brown, D. C. (1966). Decentering distortion of lenses. *Photogrammetric Engineering and Remote Sensing*, *32*(3), 444–462.

Corke, P. I., Roberts, J. M., & Winstanley, G. J. (1998). Robotics for the mining industry. In A. T. de Almeida, & O. Khatib (Eds.), *Autonomous robotic systems* (pp. 163–181). Springer.

Dahl, O. (1994). Path-constrained robot control with limited torques-experimental evaluation. *IEEE Transactions on Robotics and Automation*, *10*(5), 658–669.

Dahl, O., & Nielsen, L. (1990). Torque-limited path following by online trajectory time scaling. *IEEE Transactions on Robotics and Automation*, *6*(5), 554–561.

Duff, E., Caris, C., Bonchis, A., Taylor, K., Gunn, C., & Adcock, M. (2010). The development of a telerobotic rock breaker. In A. Howard, K. Iagnemma, & A. Kelly (Eds.), *Field and service robotics*. –. : Springer.

Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, *24*(6), 381–395.

Forsyth, D. A., & Ponce, J. (2002). *Computer vision: a modern approach*. Prentice Hall Professional Technical Reference.

Fox, J., Castano, R., & Anderson, R. C. (2002). Onboard autonomous rock shape analysis for mars rovers. In *Proceedings, IEEE Aerospace Conference* (Vol. 5, pp. 5–2052).

Girshick, R. (2015). Fast R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 1440–1448).

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 580–587).

Haddadin, S., De Luca, A., & Albu-Schäffer, A. (2017). Robot collisions: A survey on detection, isolation, and identification. *IEEE Transactions on Robotics*, *33*(6), 1292–1312.

Havoutis, I., & Calinon, S. (2019). Learning from demonstration for semi-autonomous teleoperation. *Autonomous Robots*, *43*(3), 713–726.

He, Q., Wang, Z., Zeng, H., Zeng, Y., Liu, S., & Zeng, B. (2020). SVGA-Net: Sparse voxel-graph attention network for 3D object detection from point clouds. *arXiv*, 2006.04043.

Ho, N. (2013). Finding optimal rotation and translation between corresponding 3D points. http://nghiaho.com/?page_id=671. Accessed May 16, 2021.

Horn, B. K. (1987). Closed-form solution of absolute orientation using unit quaternions. *JOSA A*, *4*(4), 629–642.

Horn, B. K., Hilden, H. M., & Negahdaripour, S. (1988). Closed-form solution of absolute orientation using orthonormal matrices. *JOSA A*, *5*(7), 1127–1135.

Hubert, G., Dirdjosuwondo, S., Plaisance, R., & Thomas, L. (2000). Teleoperation at freeport to reduce wet muck hazards. *MassMin 2000*, 173–179.

Hulttinen, L. (2017). *Position-based impedance control of a hydraulic rock breaker boom* (Master's thesis). Tampere University of Technology.

Hustrulid, W. A., & Bullock, R. L. (2001). *Underground mining methods: Engineering fundamentals and international case studies*. Society for Mining, Metallurgy, and Exploration (SME).

Jazar, R. N. (2010). *Theory of applied robotics: kinematics, dynamics, and control* (2nd ed.). Springer Publishing Company.

Koivumäki, J., & Mattila, J. (2015). Stability-guaranteed force-sensorless contact force/motion control of heavy-duty hydraulic manipulators. *IEEE Transactions on Robotics*. *31*(4), 918–935.

Koivumäki, J., & Mattila, J. (2017). Stability-guaranteed impedance control of hydraulic robotic manipulators. *IEEE/ASME Transactions on Mechatronics*, *22*(2), 601–612.

Ku, J., Mozifian, M., Lee, J., Harakeh, A., & Waslander, S. L. (2018). Joint 3d proposal generation and object detection from view aggregation. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE (pp. 1–8).

Lampinen, S., Koivumäki, J., Mattila, J., & Niemi, J. (2019). Model-based control of a pressure-compensated directional valve with significant dead-zone. In *ASME/BATH 2019 Symposium on Fluid Power and Motion Control*. American Society of Mechanical Engineers Digital Collection.

Lampinen, S., Niemi, J., & Mattila, J. (2020). Flow-bounded trajectory-scaling algorithm for hydraulic robotic manipulators. In *2020 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*.

Liang, M., Yang, B., Chen, Y., Hu, R., & Urtasun, R. (2019). Multi-task multi-sensor fusion for 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Liang, M., Yang, B., Wang, S., & Urtasun, R. (2018). Deep continuous fusion for multi-sensor 3d object detection. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 641–656).

Lovgren, R. (2004). Radical improvements in crane safety. *ISO Focus*, *1*(7), 21–93.

Manner, J., Mörk, A., & Englund, M. (2019). Comparing forwarder boom-control systems based on an automatically recorded follow-up dataset. *Silva Fenn*, *53*, 10161.

Mattila, J., Koivumäki, J., Caldwell, D. G., & Semini, C. (2017). A survey on control of hydraulic robotic manipulators with projection to future trends. *IEEE/ASME Transactions on Mechatronics*, *22*(2), 669–680.

McKinnon, C., & Marshall, J. A. (2014). Automatic identification of large fragments in a pile of broken rock using a time-of-flight camera. *IEEE Transactions on Automation Science and Engineering*, *11*(3), 935–942.

Metso Mining and Construction. (2015). Basics in Minerals Processing Handbook.

Niu, L., Aref, M. M., & Mattila, J. (2018). Clustering analysis for secondary breaking using a low-cost time-of-flight camera. In *2018 Ninth International Conference on Intelligent Control and Information Processing (ICICIP)*. IEEE (pp. 318–324).

Niu, L., Chen, K., Jia, K., & Mattila, J. (2019). Efficient 3d visual perception for robotic rock breaking. In *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE)*. IEEE (pp. 1124–1130).

Nurmi, J., & Mattila, J. (2017). Automated feed-forward learning for pressure-compensated mobile hydraulic valves with significant dead-zone. In *Proceedings of ASME/BATH Symposium on Fluid Power and Motion Control (FPMC2017)*, Oct. 16–19, 2017, Sarasota, FL.

Qi, C. R., Litany, O., He, K., & Guibas, L. J. (2019). Deep hough voting for 3d object detection in point clouds. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 9277–9286).

Qi, C. R., Liu, W., Wu, C., Su, H., & Guibas, L. J. (2018). Frustum pointnets for 3d object detection from rgb-d data. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (pp. 918–927).

Redmon, J. (2018). Darknet: Open source neural networks in c. Pjreddie.com.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (pp. 779–788).

Redmon, J., & Farhadi, A. (2017). Yolo9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7263–7271).

Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. *arXiv*, 1804.02767.

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems* (pp. 91–99).

Sandvik Mining and Construction. (2016). *Hydraulic Hammer Rammer 2577, Operator's Manual*. Available at https://www.rammer.com/en/products/hydraulic-hammers/excellence-line/medium-range/2577/

Scharstein, D., & Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, *47*(1-3), 7–42.

Sciavicco, L., Siciliano, B., & Sciavicco, B. (2000). *Modelling and control of robot manipulators* (2nd ed.). Springer-Verlag.

Smirnov, S., Georgiev, M., & Gotchev, A. (2015). Comparison of cost aggregation techniques for free-viewpoint image interpolation based on plane sweeping. In *Ninth International Workshop on Video Processing and Quality Metrics for Consumer Electronics*.

Suomalainen, M., Koivumäki, J., Lampinen, S., Mattila, J., & Kyrki, V. (2018). Learning from demonstration for hydraulic manipulators. In *IEEE/RSJ International Conference on Intelligent Robots and System (IROS)*.

Tafazoli, S., Salcudean, S. E., Hashtrudi-Zaad, K., & Lawrence, P. D. (2002). Impedance control of a teleoperated excavator. *IEEE Transactions on Control Systems Technology*. *10*(3), 355–367.

Takahashi, H., & Monden, T. (1999). Automatic system for boulder breakage using force sensors. *Fragblast*, *3*(3), 267–277.

Takahashi, H., & Sano, K. (1998). Automatic detection and breaking system for boulders by use of ccd camera and laser pointer. *Fragblast*, *2*(4), 397–414.

Walker, M. W., Shao, L., & Volz, R. A. (1991). Estimating 3-d location parameters using dual number quaternions. *CVGIP: Image understanding*, *54*(3), 358–367.

Wanner, J., & Sawodny, O. (2019). Tool-center-point control of a flexible link concrete pump with hydraulic limitations using quadratic programming. In *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE)*, (pp. 561–566).

Yang, B., Luo, W., & Urtasun, R. (2018). Pixor: Real-time 3d object detection from point clouds. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(11), 1330–1334.

Zhang, Z. -X. (2016). Chapter 21- optimum fragmentation, *Rock Fracture and Blasting* (pp. 411–423). Butterworth-Heinemann.

Zhao, X., Liu, Z., Hu, R., & Huang, K. (2019). 3d object detection using scale invariant and feature reweighting networks. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 33, pp. 9267–9274).

Zhou, Y., & Tuzel, O. (2018). Voxelnet: End-to-end learning for point cloud based 3d object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4490–4499).

How to cite this article: Lampinen, S., Niu, L., Hulttinen, L., Niemi, J., & Mattila, J. (2021). Autonomous robotic rock breaking using a real-time 3D visual perception system. *J Field Robotics*, 38, 980–1006. https://doi.org/10.1002/rob.22022