Autonomous Excavator for Precise Earthcutting and Onboard Landscape Inspection

Inkyu Jang[1*], Junha Kim[2*], Dongjae Lee[1*], Changhyeon Kim[3*], Changsuk Oh[4], Youngbum Kim[5],
Sangwook Woo[5], Heejee Sung[5], and H. Jin Kim[1]

Abstract—**Autonomous excavator systems can alleviate the issues caused by the shortage of skilled labor forces and increasing labor costs. For autonomous excavation, real-time landscape estimation, excavation path generation, control, and precise landscape inspection are all essential. In this article, we propose and experimentally validate an integrated autonomous excavator system incorporating all these elements. Specifically, unlike previous research, we introduce a sensor arrangement capable of sufficiently covering the regions of interest regardless of the inclination of the target landscape, a motion planning method that satisfies geometric and physical constraints, and a precise post-excavation inspection module using only onboard sensors. The proposed methodology was experimentally validated using a real 30-ton hydraulic excavator. It successfully performed a cutting task on an upward slope with $45°$ inclination and achieved a centimeter-level accuracy through autonomous repetitive excavation; also, the proposed post-excavation inspection method demonstrated sub-centimeter precision within seconds using onboard sensors only.**

Index Terms—**Hydraulic excavator, automation in construction, motion planning, landscape inspection, computer vision.**

I. INTRODUCTION

THE need for automation in construction has been consistently growing due to rising labor cost, lack of skilled laborers, and stricter safety requirements. This has led to active research both in the industry and the academia, including autonomous driving for heavy equipment vehicles [1], autonomous trenching for hydraulic excavators [2]–[4], dry stone wall construction [5], and semi-automation in wheel loader material loading [6]. Specifically, tasks like large-scale land cuttings and precise surface grading, essential at the beginning of most construction sites, can benefit from automation as they

Fig. 1: The 30-ton hydraulic excavator hardware considered in this research. The proposed AES is capable of performing earthcutting and post-excavation inspection works fully autonomously using only onboard sensors.

involve repetitive actions requiring precision and accuracy. These operations typically consist of two subtasks: earthcutting using excavators and post-excavation topographic inspection via external equipment. Traditionally, both are carried out by skilled human workers who must sometimes risk their safety in hazardous workspaces full of massive equipment. A fully onboard integrated autonomous excavator system (AES) which reduces the need for external equipment and human effort, therefore, possesses huge potential and advantages.

The requirements for the first subtask, earthcutting, are multifaceted. First, a decision-making module that determines where to dig and generates the path for the arm part are needed. This module should be capable of creating a digging profile and the arm path within seconds to ensure uninterrupted workflow. Secondly, to facilitate the aforementioned path planning, a perception functionality providing information about the current landscape is necessary. This perception functionality needs not be overly precise considering the size and control accuracy of the hardware, and it should be capable
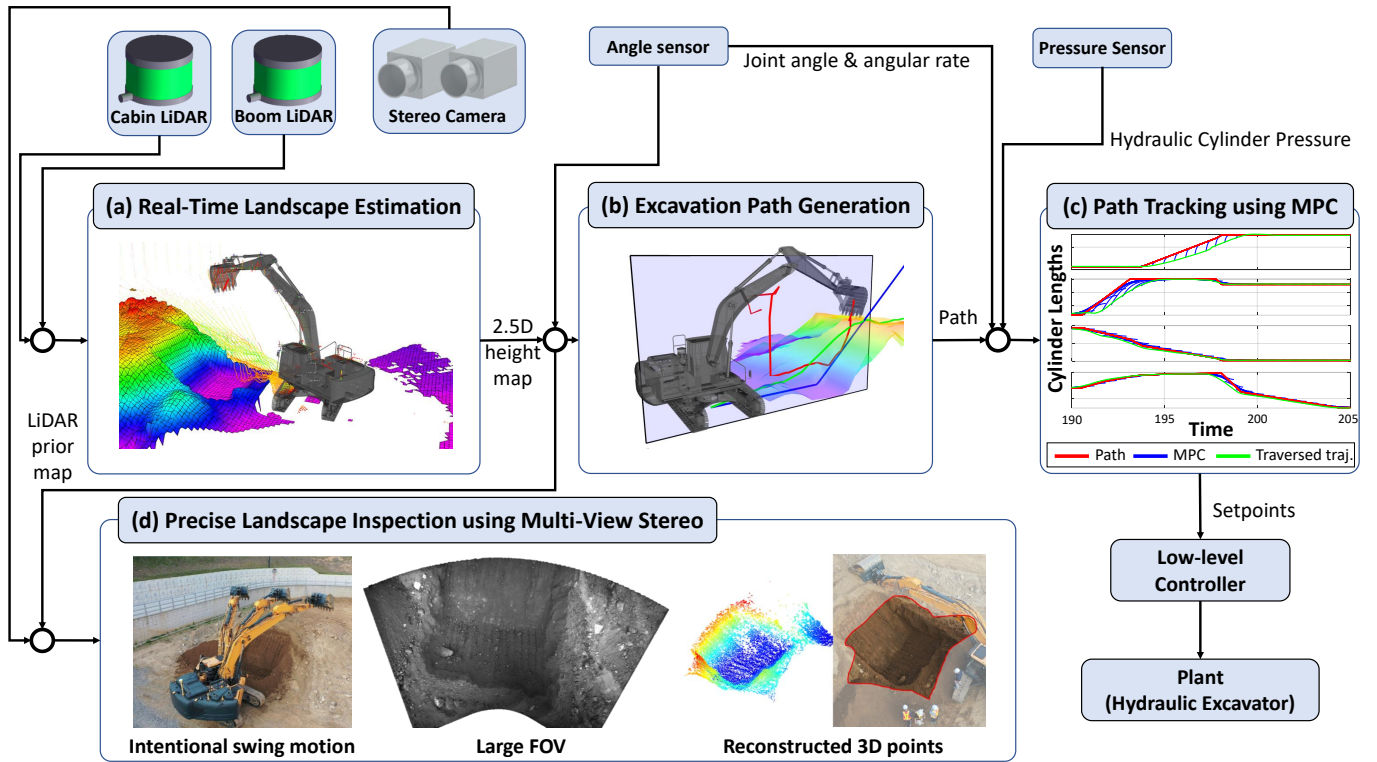
Fig. 2: The overall flowchart of the proposed AES. Its software components include (a) the real-time landscape estimation module, the motion planning module with (b) excavation path generation and (c) path tracking functionalities, and (d) the precise landscape inspection module.

of capturing real-time changes in the terrain made by the digging and dumping actions. Lastly, to track the generated path despite disturbing forces such as bucket-soil interaction, a local trajectory planner offering real-time tracking should be in place. This local planner should take into account the physical limitations of the excavator, such as power and arm speed constraints.

The second subtask, post-excavation landscape inspection, aims to verify whether the targeted terrain has been accurately shaped. This subtask is conducted separately from earthcutting, in order to ensure the integrity of the assessment process and to prevent any potential influence from earthcutting motion. Construction sites typically employ specialized surveying personnel, who submit the results as evidence of completion. Although construction standards may vary by region and the type of task, a typical goal is to achieve construction accuracy within a few centimeters of the target terrain. To meet this requirement, the inspection module should be able to perceive the landscape with an accuracy level of less than 1 cm. This should be carried out within seconds to prevent delay in the construction process.

Despite several attempts both in the industry and the academia to implement a fully-autonomous excavator, none of the existing solutions have performed both motion generation and post-excavation inspection onboard [3]–[5]. In this research, we present an AES (Fig. 1) that combines real-time landscape estimation, onboard motion generation, and accurate post-excavation topography inspection–all achieved solely through onboard apparatus. Our system improves existing AES

solutions in the following perspectives:

- **Sensor configuration**: Our sensor configuration can capture the region of interest (ROI) within their field of view (FOV) regardless of the specific task at hand. This sensor configuration employs two 3D light detection and ranging (LiDAR) sensors mounted at different parts of the excavator. A stereo camera attached at the bottom of the boom provides a sufficiently large FOV. Thanks to this configuration, we can measure the ROI of the workspace more densely on any terrain slope.
- **Landscape estimation**: The proposed AES is capable of onboard-performing real-time estimation of the surrounding landscape, which facilitates digging profile generation and arm-ground collision avoidance. The proposed Kalman-filter-like update rule provides a reliable estimate of the environment, even during interaction between the bucket and the ground.
- **Constraint-satisfying motion generation**: We propose hierarchical motion planning software that allows the trajectory of the excavator arm to adhere to both geometric (related to the shape of the excavation profile) and physical (related to the hardware's actuation limits) constraints. A high-level grid-based path planning algorithm running faster than 1 Hz computes an excavation path considering the geometric constraints, and a low-level real-time MPC running at 10 Hz generates a time-allocated trajectory that tracks the excavation path while satisfying the physical constraints.
- **Post-excavation inspection**: Our AES is also equipped

with an onboard landscape inspection functionality that uses a stereo camera. This enables verification of the work quality without the need for additional equipment or workers. Making the best use of the excavator's configuration and motion, it is capable of examining the completed landscape within a few seconds at sub-centimeter-level accuracy even under severe vibration and impact caused by the excavation work. To the best of the authors' knowledge, it is the first autonomous heavy equipment with an onboard functionality of post-excavation landscape inspection, distinct from real-time landscape estimation during excavation.

We demonstrate the performance of the proposed integrated AES with actual hardware in a $45°$ upward slope-cutting scenario. The proposed AES successfully shaped the desired landscape and conducted a post-excavation topography inspection with a mean error of less than one centimeter.

II. RELATED WORK

Fully-autonomous excavators are receiving considerable attention from both industry and academia. In the industry, Built Robotics provides an autonomous safety-aware trenching solution incorporating sensors and software [2]. In the academia, [3] provides an integrated framework including sensing, planning, and control for a fully autonomous excavator, and shows applicability in multiple tasks involving trenching, material loading, and rock removal. Similar tasks were also performed using a walking excavator [4], [5] in a fully autonomous manner. However, none of the attempts tackled integration of earthcutting and post-excavation topographic inspection. In the remainder of this section, we introduce related works in the perspectives of excavator sensor configuration, environment perception, and excavation motion generation.

Thanks to their easy usage and reliability of measurement data, 3D LiDAR sensors have been widely employed for autonomous excavators. In deploying them, existing works on system integration [3], [5], [7]–[9] installed one or two 3D LiDAR(s) on top of the cabin. However, in the case of slope cutting, such configuration may have limited FOV, and the ROI for excavation may not be fully covered. In [5], [10], a drone with a camera was used, but additional burden of operating these equipment naturally follows. To achieve self-contained autonomous excavation, a strategy that ensures a large FOV without external equipment is desirable.

In a typical construction site, landscape inspection should be conducted separately after the earthcutting work, because it serves as confirmation of completion of the task. Although the actual value varies depending on the type of construction work, inspection should be done at an accuracy of less than a centimeter. While this precision is typically achieved by using land surveying solutions [11], their drawback lies in the need for separate workers and equipment. Furthermore, these solutions are expensive and can take tens of minutes, even for scanning a small area up to 10 m^2. Existing autonomous heavy equipment [3], [5], [7]–[9] utilized 3D LiDAR sensors to perform a perception of the surrounding landscape, but whether they are accurate and reliable enough to replace the inspection process is still unclear, as there was little analysis on that.

Excavation trajectory should abide by both 1) geometric constraints such as vehicle-ground collision avoidance, and 2) physical constraints including power and flow rate limits, as violation of any of these may lead to dangerous situations such as tripping over or failure of accomplishing the given task. Although many studies considered the motion constraints, they either admitted constraint violation [4], [7], [12] or partially omitted the physical constraints [13], [14]. Only a few of them such as [15], [16] addressed both types of motion constraints; unfortunately, they were tested only in simulation environments. Their robustness against noisy data and real-time applicability (in terms of computation time), therefore, were not validated with real-world hardware.

III. SYSTEM OVERVIEW

The proposed AES is capable of completing an *excavation task* without human intervention, where we refer a single excavation task to the whole process of achieving the desired topography, after the base part (crawler or wheels) has settled to a designated place, and before it leaves from the location. More specifically, an excavation task in our AES consists of the following process.

1) The excavator repeats dig cycles to make the landscape resemble the desired topography.
2) If the current landscape becomes sufficiently close to the desired one, a grading process using the bucket tip is carried out for precise landscape shaping.
3) Using onboard sensors only, the excavator conducts precise landscape inspection.
4) If the inspection result has reached satisfactory precision and accuracy, the task terminates.

To successfully complete such an excavation task, we design the proposed AES to have the following four features:

- uncrewed-controllable hydraulic excavator hardware equipped with an onboard computer and sensors,
- a real-time landscape estimator providing information about the surrounding environment,
- a motion planning algorithm generating and tracking the motion of the excavator arm, and
- post-excavation landscape inspection software based on computer vision.

During the task, the real-time landscape estimation module consistently provides real-time estimates of the current landscape, given information from the onboard sensors attached to positions carefully selected for better perception capabilities. Until the real-time observed landscape sufficiently resembles the desired landscape, the motion planning module iteratively computes an excavation path and a trajectory that tracks the generated path. Then, the precise inspection software examines the completed topography after the grading process. Fig. 2 provides a brief overview of how the four features collaborate to complete each step of the excavation task. In the following four sections, we delineate the details of each feature.

(a) The hardware components of the proposed AES

(b) Sensor configuration

(c) 3D LiDAR pointcloud in the simulation environment

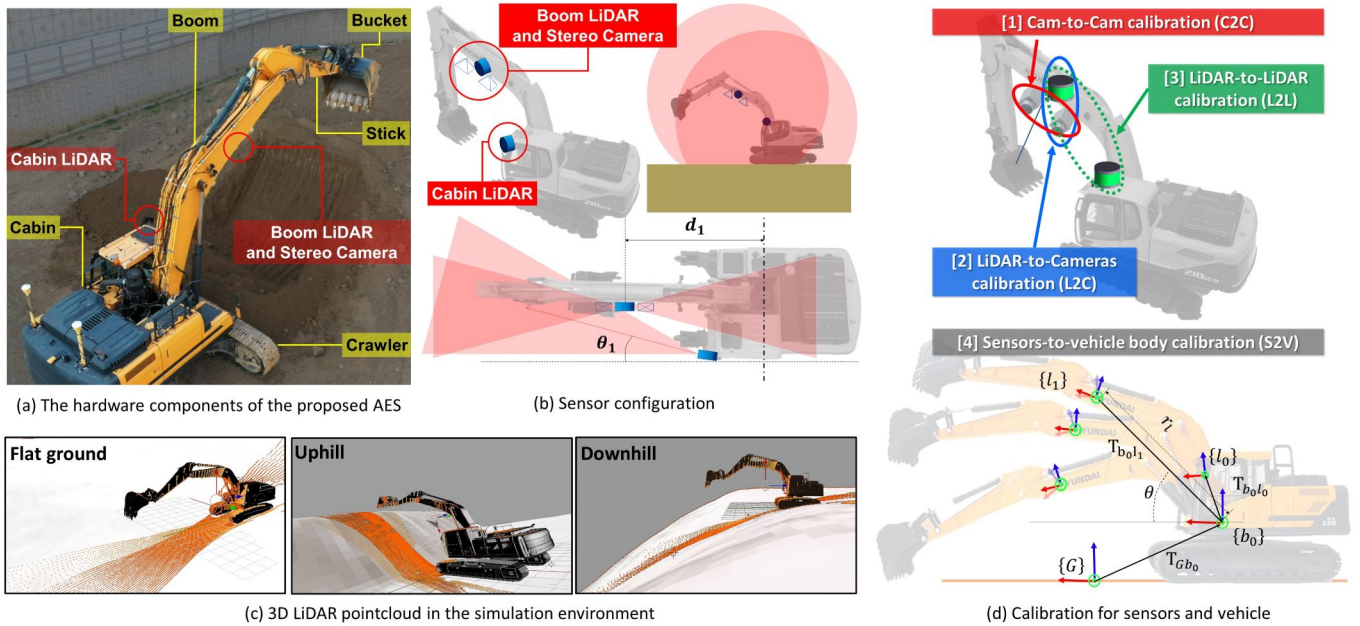(d) Calibration for sensors and vehicle

Fig. 3: (a) The hardware components of the proposed AES consist of an unmanned-controllable hydraulic excavator with onboard cameras and 3D LiDAR sensors. The excavator kinematics can be understood as a robotic manipulator with five joints (crawler, cabin, boom, stick and bucket) connected in series through revolute joints. A stereo camera is attached on the boom, and two 3D LiDARs are attached on the cabin and the boom, respectively. The yellow labels denote the excavator parts, while the red ones represent the sensors. (b) The configuration of the employed sensors: two 3D LiDARs and two cameras. Sensor configuration is carefully selected to secure a uniformly and sufficiently large FOV in various arm (boom, stick, and bucket) configurations. This is to always include the ROI inside the FOV during slope cutting with arbitrary inclination, as shown in (c). (d) Sensor calibration. The numbering indicates the order of the calibration. Thanks to the carefully selected sensor configuration, sensor calibration can be performed with only one or two human worker(s), since the FOVs of the onboard sensors overlap in the ROI on the ground, which is easily accessible without additional equipment such as ladders. In addition, for the whole calibration process, a checkerboard is only required as the calibration target.

IV. HARDWARE

A. Hydraulic Excavator

As shown in Fig. 3 (a), the kinematic model of the excavator consists of five components: crawler, cabin attached to the crawler with a vertical revolute joint, and the arm part composed of three serially-connected rigid bodies (boom, stick, and bucket). We call the rotation of the cabin with respect to the crawler the swing motion. Four hydraulic joints actuating swing, boom, stick, and bucket can be autonomously manipulated using signals relayed from an onboard computer.

B. Sensor Configuration

To obtain kinematic information, inertial measurement units (IMUs) mounted on the cabin, boom, stick, and bucket measure the configuration of the arm part. A rotary encoder in the revolute joint connecting the crawler and the cabin measures the swing angle. Angle measurements are utilized in excavation path generation and control, while cylinder pressures measured by sensors inside each hydraulic cylinder are employed in defining a power constraint in motion planning.

As shown in Fig. 3 (b), two 3D LiDARs and a stereo camera are used. To ensure large FOV during slope cutting, we attach one 3D LiDAR (i.e., *boom LiDAR*) at the bottom of the boom where the rotating axis of the boom LiDAR is parallel to the rotating axis of the boom. Since the boom LiDAR moves along with the boom, dense pointcloud data in the ROI can be obtained for landscapes with any inclination as shown in Fig. 3 (c). This allows the excavator to well estimate the

surrounding landscape during excavation without the need for an intentional boom motion.

To additionally provide consistent observation of the ROI regardless of the motion of the boom, we attach the other 3D LiDAR (i.e., *cabin LiDAR*) to the side of the cabin with a slight tilt angle. The cabin LiDAR compensates the boom LiDAR, and the landscape can be measured in the distance without severe occlusion by the arm part. By using cabin and boom LiDARs together, a denser pointcloud at the ROI can be obtained.

To obtain precise and accurate environment reconstruction during landscape inspection, we utilize the stereo camera with the boom and cabin LiDARs. A stereo camera attached at the bottom of the boom alongside the boom LiDAR provides a sufficiently large FOV, regardless of the inclination of the landscape the excavator is facing. This is because, during inspection, the boom angle no longer needs to be manipulated for excavation but only for inspection. As we discuss in the remaining sections, the use of LiDAR data in environment reconstruction provides good prior knowledge of the landscape and therefore enhances computation speed.

C. Sensor Calibration

Note that all the sensors' FOVs overlap near the ROI which is easily accessible by human workers without the need for a special equipment such as a ladder, as shown in Fig. 3 (c). This allows the calibration of the attached sensors to be done efficiently by only one or two workers, thereby reducing human effort, especially considering that the arm reach of the

equipment typically ranges up to 10 m. The whole calibration process only requires a single checkerboard as the target. Using a checkerboard, we estimate the intrinsic parameters of cameras and extrinsic parameters between sensors, and between each sensor and the excavator: camera-to-camera (C2C, i.e., within the stereo camera), LiDAR-to-camera (L2C), LiDAR-to-LiDAR (L2L), and sensors-to-vehicle (S2V) calibration, as shown in Fig. 3 (d).

C2C and L2L calibrations are performed using [17] and [18], respectively. The extrinsic parameters between the 3D LiDARs and the stereo camera are estimated by the automated calibration target detection algorithm proposed in [18] and solving an optimization problem to minimize the distance between target planes estimated by each sensor measurement in the same epoch. Then, through L2L calibration, we calculate the relative pose between the two 3D LiDARs for various boom angles. With the obtained relative pose, we solve for boom angle (θ) at each epoch and the distance (r_l) from the boom rotation axis to the boom LiDAR. Finally, the S2V extrinsic calibration is performed by using poses of cabin LiDAR and boom LiDAR with respect to the frame $\{b_0\}$ on the boom's rotation axis. As a result, relative poses between all sensors and the vehicle can be calculated for any boom angle measured by IMU attached to the boom.

V. REAL-TIME LANDSCAPE ESTIMATION

For real-time landscape estimation, we maintain a grid-based 2.5D elevation map and update it in real-time using 3D LiDAR sensor measurements similar to [7]. Such estimation inheres the following challenges:

- Undesirable 3D LiDAR points (not from the landscape) can be observed by various disturbing factors including dust in construction, soil falling from the bucket, and the excavator itself.
- The shape of the ground may change through the use of a bucket (excavation work) and even when there is no excavation, due to spontaneous micro-scale landslips caused by local slope.

To address the first challenge, we first filter out body-occluded points for every incoming 3D LiDAR pointcloud data. This rejection is done by considering the kinematics and geometric model of the excavator: As shown in Fig. 4 (c), any 3D LiDAR point falling inside any one of the box-shaped regions is rejected.

Although the proposed point rejection method sorts out most of the undesired points, there may still exist many non-rejected points. They originate mainly from the mechanical vibration of the sensor base or parametric error between the geometric model and the actual excavator. Also, during dumping, the falling dirt would produce undesirable 3D Li-DAR points which the above method cannot properly reject. Moreover, the deformation of the landscape itself consistently occurs throughout the excavation task, necessitating a separate technique that is capable of capturing landscape changes in real-time.

To address the concerns, we assume that the landscape follows a certain stochastic *dynamics* model, and run a Kalman-like filter based on this dynamics. The whole 2.5D elevation map is considered as a dynamic system, with the elevation of the cells being its state. Let us denote the elevation of cell c at time t_k by x_k^c. Then, x_{k+1}^c (the elevation of the cell at the next time epoch t_{k+1}) is determined by the following update rule.

$$x_{k+1}^c = \begin{cases} x_{\text{bucket}}^c + u_k^c & \text{(A)} \\ x_k^c + w_k^c & \text{(otherwise),} \end{cases} \tag{1}$$

where the update case (A) occurs if the bucket tip has swept below the ground level at elevation x_{bucket}^c on the cell between time t_k and t_{k+1}, and u_k^c, w_k^c are white Gaussian noises with variance Q_u and Q_w, respectively.

Since the noises u_k^c and w_k^c represent the discrepancy between the model and actual elevation level, the values of Q_u and Q_w can be understood as measures indicating how unpredictable the landscape changes with respect to time. For example, with small Q_u and Q_w values, the estimator believes that the landscape tends to remain still, and with larger values, it assumes that changes in the landscape (e.g., micro-scale landslips) occur more frequently, thereby relying more on recent measurements given from the LiDAR sensors. In case the actual values of noise cannot be directly measured, the values of Q_u and Q_w can be estimated statistically.

For filtering, each cell is assigned not only the estimated expectation of its elevation but also the estimation variance. Initially, every cell is assigned infinite variance and zero expectation, although the expectation has no significance under infinite variance. Let a non-rejected 3D LiDAR point be included in cell c with elevation y. Since one can presume that a 3D LiDAR point closer to the center of the cell will better represent the true elevation at the cell (i.e., the elevation at the center of the cell), we assume that y follows the model

$$y \sim \mathcal{N}(x_k^c, (1 + C \cdot d^2) \cdot v), \tag{2}$$

where $v > 0$ is the variance of the distance measurement from 3D LiDAR (given from the sensor specs), d is the horizontal distance from the 3D LiDAR point to the center of the cell c, and $C > 0$ is a parameter related to the maximum slope of the excavating landscape. We obtain a maximum a posteriori (MAP) estimate of the elevation of every cell at every update. Fusion for mean and variance is conducted using the elevation in the previous timestep t_k, and the aggregation of incoming 3D LiDAR pointclouds between t_k and t_{k+1}. Such computed MAP mean, $\hat{x}_{k+1}^c$, represents the topographic information at time t_{k+1} and is employed for collision avoidance and excavation path generation.

Empirically, we found that the performance of the proposed landscape estimator does not get significantly affected by the values of the tunable parameters, such as the size of the cell, Q_u, Q_w, C, and the time window $\delta t = t_{k+1} - t_k$, as long as their values remain reasonable. It can be seen in Fig. 4 (a), (b) that the proposed landscape estimator succeeds in stably capturing the change of the ground shape, even while (a) digging and (b) dumping where reliable 3D LiDAR measurements are not available due to the bucket movement and the falling dirt. Snapshots of landscape estimates taken during an excavation task are shown in Fig. 4 (d). As the
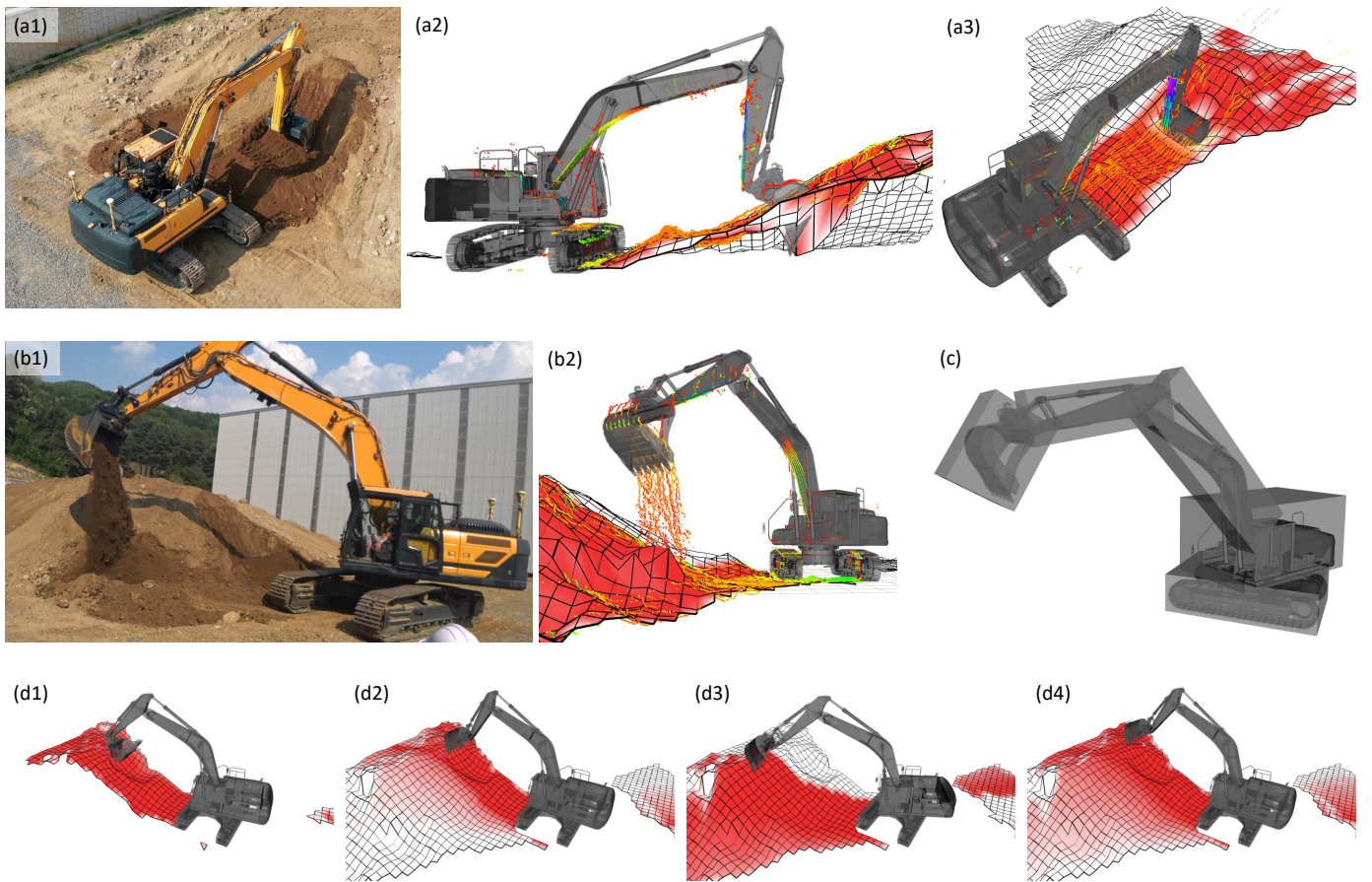
Fig. 4: Real-time landscape estimation. Subfigures (a) and (b) depict the proposed landscape estimation module in operation during digging and dumping, respectively. In both situations, even though reliable sensor measurements are not available due to the bucket motion and the falling dirt, the proposed estimator successfully captures landscape changes in a stable manner. This is possible due to the kinematics-based pointcloud filtering (any point falling into one of the boxes shown in (c) is rejected), and the Kalman-like MAP estimation method assuming a stochastic dynamics (1) of the landscape. As multiple swing motions are made while excavation progresses, the 3D LiDAR measurements naturally reveal additional information about the surrounding landscape, as shown in (d). Throughout this figure, the color of the grid cell represents the covariance of the estimated elevation; red and white represent small (high estimation confidence) and big (low confidence) covariance, respectively. Grid cells with infinite covariance, i.e., never-observed cells, are not shown.

excavation task progresses, the 3D LiDAR measurements reveal more area around the equipment.

VI. EXCAVATION MOTION GENERATION AND TRACKING

To successfully accomplish an excavation task, generated motion has to satisfy both geometric constraints (e.g., collision avoidance while not in excavation, or bucket pose constraint while digging) and physical constraints (e.g., constraints on hydraulic cylinder speed, pump flow rate, and power limit). Since geometric constraints can be formulated in configuration space, they can be imposed on a *path* without time information, and offline computation of such path would suffice. On the other hand, because physical constraints require real-time state and input feedback and are defined on the entire state space, the constraints should be enforced on a *trajectory*, and the trajectory should be calculated online in real-time. To fulfill all these constraints, similar to our previous work [15], we propose a two-staged motion generation scheme composed of path generation and path tracking motion generation.

A. Excavation Path Generation

The path generation functionality is mainly devoted to satisfying the geometric constraints. In generating an excavation path, we consider a number of geometric constraints as follows. First, during digging,

- The swing angle should be fixed, i.e., the motion of the arm part should be confined on the *excavation plane*, the plane spanned by the workspace of the arm while the swing angle is fixed (Fig. 5 (a)).
- To avoid over-excavation, the bucket should always stay above the target landscape.
- To avoid damage to the joints, excavator parts other than the bucket should not touch the ground.

And while not in digging,

- Each part should not collide with the ground.
- When transporting a bucket filled with dirt to the dumping site, the bucket orientation has to be fixed not to spill the dirt.

To effectively compute a path that completes an excavation task while satisfying all the above geometric conditions, we first divide a single digging task into five phases as illustrated in Fig. 5 (b):
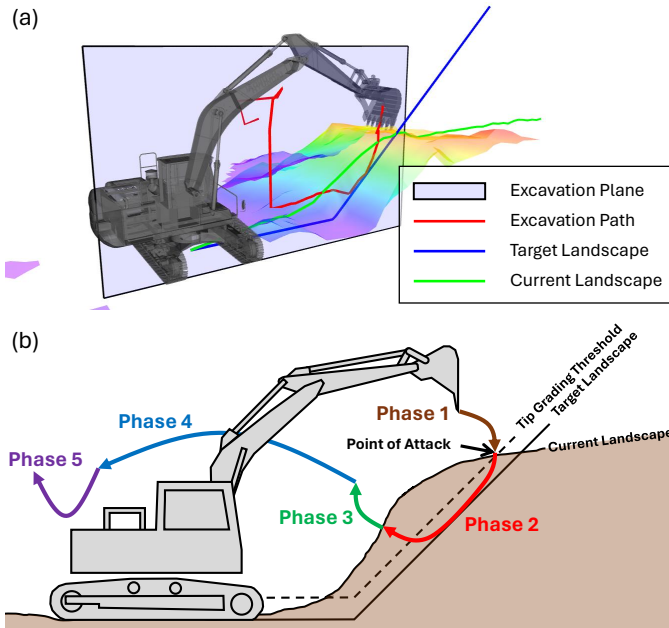
Fig. 5: (a) An example of the excavation path generated during the experiment. The *excavation plane* is shown as a blue transparent box. It is defined as the plane spanned by the workspace of the arm, while the swing angle being fixed with the cabin pointing at the desired digging direction. (b) The digging path consists of five phases. Phase 1: The bucket moves to the *point of attack*. Phase 2: The bucket cuts through the ground. Phase 3: The bucket is lifted above the ground. Phase 4: The dirt inside the bucket is conveyed to the dumping position. Phase 5: Dumping occurs.

Phase 1: The bucket moves from the initial configuration to the *point of attack*.

Phase 2: Digging occurs.

Phase 3: The bucket filled with dirt is raised above the ground.

Phase 4: The excavator conveys the bucket to the dumping position, while not spilling the dirt.

Phase 5: Dumping occurs.

To ensure path feasibility and continuity across phases, we first compute the path for phase 2, i.e., the digging profile, and then for the other phases.

For efficient and precise earth shaping, we establish two different types of digging profiles: normal digging, and tip grading. The normal digging profile is intended to scoop the largest possible amount of soil every repetition while not exceeding the bucket's capacity, so that the number of digging-dumping cycles is reduced. On the other hand, the tip grading profile aims to precisely shape the ground as desired while not penetrating through the target landscape. The normal digging profile is first employed, and then the tip grading profile is used after the current landscape approaches the target landscape sufficiently close after some digging repetitions.

To establish the criterion on how to switch between the two digging profiles, we first define three ground shapes as lines on the excavation plane: the current landscape obtained from real-time landscape estimation module, the prescribed target landscape, and the tip grading threshold located above the target landscape by a certain threshold, as shown in Fig. 5 (b). The excavator selects the tip grading profile if the current ground shape lies between the target landscape and the tip
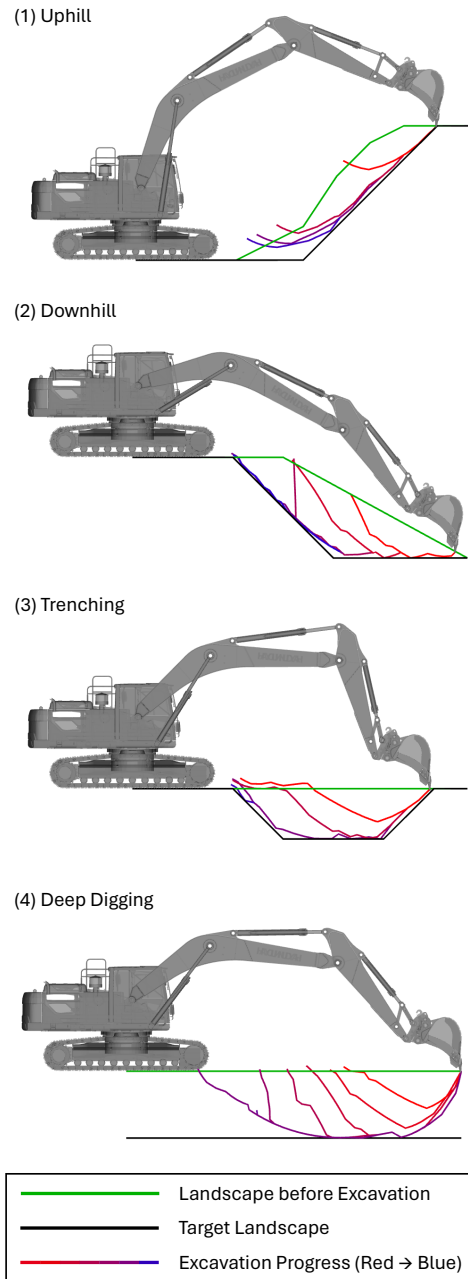

Fig. 6: Normal digging profiles were generated in a simulation environment with different target landscapes. Regardless of the current and the target ground shapes, the proposed path generation module successfully generated consecutive digging profiles without encountering infeasibilities.

grading threshold, and the normal profile otherwise. Note that due to the piling-up dirt and micro-scale landslips, it is also possible to switch back from tip grading to the normal profile.

The normal digging profile is generated with the following procedure. We assume that the target landscape is *achievable* given the given excavator hardware and its crawler position, and it has no cavity. First, the excavator attempts to find the point of attack for excavation. The point of attack is decided to be an intersection of the current landscape and the tip grading threshold with the farthest distance from the cabin, as shown in Fig. 5 (b). In order to search for a feasible digging path while dealing with the path-dependent nature of the swept volume,
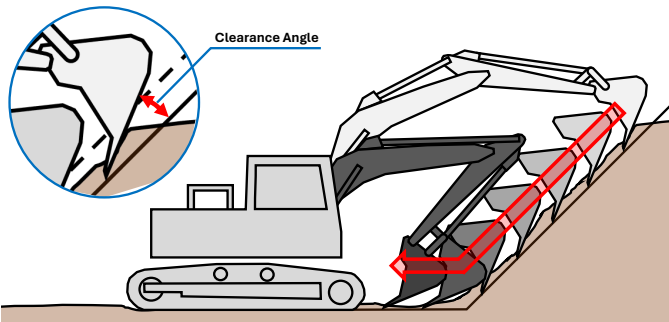
Fig. 7: An example of a tip grading profile. The bucket tip follows the desired landscape, starting from its farthest reachable point on the target. The big red arrow denotes the moving direction of the bucket. The cutting angle is kept constant throughout the profile.

we make use of a tree search algorithm on the 3D grid whose three axes consist of the lengths of boom, stick and bucket cylinder lengths, respectively. Feasibility here refers to bucket capacity, the cutting angle (the angle between the tip path and the bucket), and whether the arm penetrates through the target landscape. To reduce the number of searches, we employ a depth-first search (DFS) algorithm on the 3D grid defined by the lengths of the three hydraulic cylinders composing the excavator arm. Among the tree nodes with same depth, the priority is determined using a hardware-specific heuristic that takes into account both the expected excavation volume and the expected power consumption of the cylinders. By using DFS, we are able to find a feasible path that scoops satisfactory volume at a relatively low computational cost. Fig. 6 shows how the normal digging profile is generated in different landscapes. It can be seen from the figure that in each digging-dumping cycle, the amount of soil removed by the bucket is kept approximately constant, where the removed volume is close to the bucket's capacity.

The tip grading profile, on the other hand, is generated in a manner that the bucket follows the target landscape. An example of a tip grading profile is shown in Fig. 7. As shown in the figure, starting from the farthest reachable point on the target landscape, the bucket tip is pulled toward the cabin, while maintaining a constant cutting angle.

In phases 1, 3, 4, and 5, we apply A* planner to generate a path in a 4D grid whose four axes comprise the swing motion and three cylinder lengths. The objective of the planner is to reach the goal configuration while considering the excavator kinematics and minimizing the aggregate flow in and out of each cylinder. In phase 4, in order to prevent the excavator from spilling the removed dirt, we impose an additional constraint where a configuration space having the open direction of the bucket not heading upward is considered as an obstacle.

B. Path Tracking using Model Predictive Control

Since the generated excavation path lacks time information, we perform time allocation before execution. We define the traversal time of each path segment to be maximal but not to exceed the reference expansion/contraction velocity of each cylinder. Then, to enforce physical constraints on the time-allocated trajectory, we employ model predictive control

(MPC) in the local motion planner. MPC solves a constrained optimal control problem iteratively in a receding horizon manner and has been widely adopted in various robotics and construction applications [15], [19] since it enables to compute state and input trajectories satisfying motion constraints in real-time.

The local planner assumes the following discrete-time first-order dynamical model:

$$q_{k+1} = f_d(q_k, u_k), \tag{3}$$

where q is the 4-dimensional configuration of the excavator including the swing angle and the posture of the arm, and u is the 4-dimensional control input that comprises the swing rate and extension/retraction rate of the hydraulic cylinders. The subscript k indicates that the corresponding variable is from the time step t_k, and f_d is a function determining the next state q_{k+1} from the current state and input.

The dynamical model is derived from the first-order kinematic relation $\dot{q} = J(q)u$ where $J(q)$ is the Jacobian matrix. A better prediction model could be applied using a higher-dimensional dynamical model that incorporates hydraulic pressure dynamics and inertial effects. However, due to the high uncertainty raised by oil temperature and hydraulic friction, model identification becomes extremely challenging, and the strong nonlinearity of the high-dimensional dynamics hinders the real-time applicability of MPC. To overcome these challenges, we opted for a relatively simple kinematics-based dynamical model in MPC, leveraging a low-level controller with sufficient tracking performance. The effectiveness of this approach has been validated through experiments.

Physical constraints considered in MPC are 1) power limit, 2) cylinder displacement limit, and 3) pump flow rate limit, and can be expressed in the following form at the time step t_k:

$$c(q_k, u_k, f_p) \geq 0, \tag{4}$$

where f_p is the measurement of hydraulic pressures inside the cylinders and is used to express the power limit constraint. Due to difficulty in accurately estimating the hydraulic pressure dynamics, we choose to assume constant pressure during the prediction horizon and instead take a fast update rate. Detailed expressions for the considered physical constraints can be found in [15].

Using (3) and (4), we formulate an MPC problem as follows:

$$\begin{aligned} \min_{q_{(\cdot)}, u_{(\cdot)}} \quad & J(q_{(\cdot)}, u_{(\cdot)}; q_{r(\cdot)}) \\ \text{s.t.} \quad & q_{k+1} = f_d(q_k, u_k), \quad q_0 = q(t) \\ & c(q_k, u_k, f_p) \geq 0. \end{aligned} \tag{5}$$

We take the cost function J be quadratic with respect to the state error $q - q_r$ and input u; this is to track the high-level reference trajectory while regulating inputs so that excessively aggressive maneuvers are avoided. The computed local trajectory is then transformed to push/retract force input for the hydraulic cylinders by the low-level controller and relayed to the excavator.

The formulated MPC problem is solved using a constrained iterative linear quadratic regulator (iLQR) algorithm with the
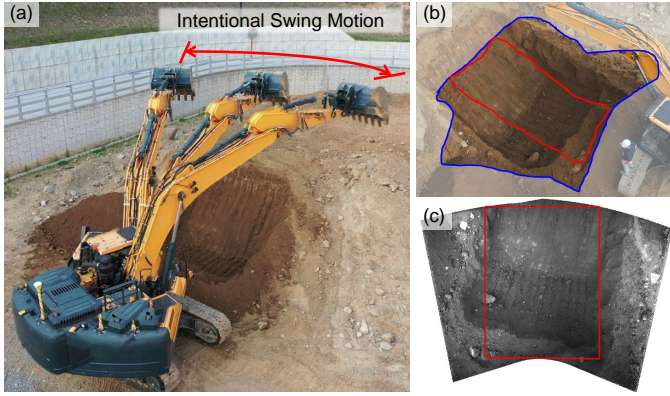
Fig. 8: (a) Intentional swing motion of the excavator. It can be seen from (b) that the proposed multi-view stereo method (blue region) not only provides a larger FOV but also enhances the accuracy and precision of landscape reconstruction, compared to the single-view counterpart (red region). (c) A stitched image generated using the BA algorithm. This image is generated using 140 single-view images and $-30° \sim 30°$ swing motion with the boom being fixed.

augmented Lagrangian method [20]. We validate its real-time applicability and constraint satisfaction in experiments. On the onboard computer, our MPC was able to run consistently at 10 Hz or faster, given the time horizon of 6 s, time discretization and integration interval of 0.04 s. Furthermore, the proposed MPC ensures to satisfy all the physical constraints as visualized in Fig. 10 (c). An ablation study for the imposed physical constraints (baseline results in the figure) also validates effectiveness of our MPC formulation.

VII. PRECISE LANDSCAPE INSPECTION USING MULTI-VIEW STEREO

The proposed 3D LiDAR-based real-time landscape estimation module captures terrain changes promptly and reliably. However, it is not suitable for post-excavation inspection, due to the following reasons. Firstly, the structural vibration corrupts the relative pose information between the two 3D LiDARs. The S2V alignment quality decays with respect to time due to extreme vibration and impact during the digging task. Secondly, the landscape estimation exhibits sensing inaccuracy of up to decimeters due to the non-negligible grid size. One might use smaller grid cells, but the inaccuracy originating from the poor LiDAR alignment still remains unresolved. Finally, the lack of 3D features on the unstructured construction site and the sparsity of 3D LiDAR pointcloud data makes pointcloud alignment hard, posing difficulty in employing LiDAR mapping methods for inspection.

We therefore construct a separate module using a stereo camera attached to the bottom of the boom for landscape inspection. Cameras provide high-resolution information at a relatively low cost, which leads to high-quality 3D reconstruction. The proposed vision-based inspection module does not require additional external apparatus and operates with onboard sensors only, completing the process within seconds. This inspection module shows a competent performance compared to land surveying solutions in precision and accuracy. Additionally, since the two cameras are rigidly connected to each other by short distance, the stereo camera is more reliable

compared to 3D LiDARs in terms of robustness against severe vibration and impact.

To ensure a sufficient amount of observation data and larger horizontal coverage of the stereo camera, intentional swing and boom motions are introduced during the inspection. Since the camera is mounted under the boom facing downward, such motions will allow the camera to view the terrain from multiple viewpoints. Fig. 8 (a) illustrates excavator motions that result in horizontal motion. Empirically, obtaining observations with swing motions of tens of degrees without a boom motion is sufficient to achieve sub-centimeter level precision.

We employ the Kanade-Lucas-Tomasi (KLT) algorithm for feature tracking. The original KLT algorithm assumes that the camera always observes the environment in a vertical direction, which is not our case. Thus, to achieve better performance in feature tracking, we model patch deformation with an affine pixel transformation and solve an optimization problem. A constrained affine KLT-based approach to track features in the target region is employed, where the initial guess is given from the real-time estimated topography. Through correspondences of feature points in sequential images, we reconstruct the 3D points of features and estimate camera poses at each sequence.

Bundle adjustment (BA) is employed, which jointly optimizes all camera poses and 3D points using multiple image measurements. Dense representation with millions of 3D points for landscape inspection yields many pose-and-point coupling terms, which may result in burdensome computation. To that end, we propose a simple but effective two-step optimization. In the first step, all camera poses and the reduced number of sampled points are jointly optimized. In most cases, we found that camera poses are accurately optimized with only thousands of points. We sample several hundred features per image with the highest Features from Accelerated Segment Test (FAST) [21] scores. In the second step, by fixing the optimized camera poses, point-only BA is executed to optimize all 3D points. By doing this, we can avoid computationally demanding calculations of pose-point coupling terms and make the proposed inspection module work within a few seconds. Note that the difference in the overall accuracy between the proposed two-step and full BA is negligible, but the two-step method is significantly faster than the full BA. A qualitative comparison of the two methods' computation time in different inspection scenarios are given in Table I.

For various terrain configurations in precision earthcutting scenarios as shown in Fig. 9, the landscape inspection module utilizing an onboard stereo camera achieves an accuracy level of 1 cm, which outperforms the two baselines using 3D LiDAR only or single-view stereo reconstruction as shown in Table II. In the table, it can also be found that the shape of the target terrain does not significantly affect accuracy. This level of accuracy surpasses the precision achievable by skilled human operators operating commonly used construction excavators of general sizes (20 tons or larger).

VIII. EXPERIMENTAL VALIDATION

A. Experiment Setup

Now, we validate the proposed AES using a real hardware in a cut slope task, in which we want the excavator to cut a

(a) Smooth slope   (b) Slope with a crater   (c) Uneven slope   (d) Flat ground
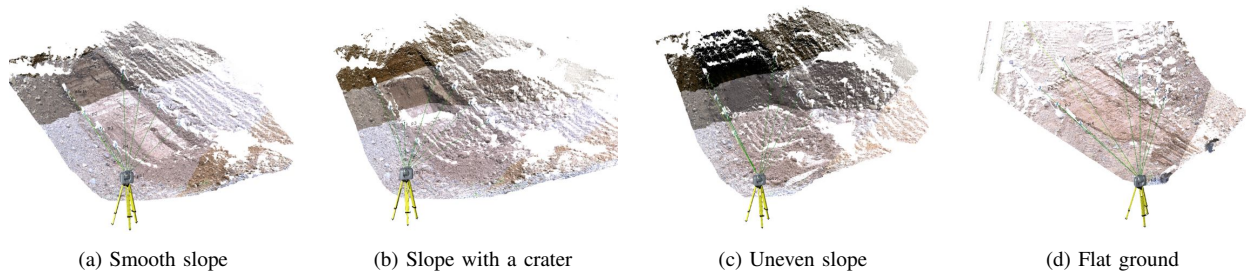
Fig. 9: Four scenarios to validate the proposed landscape inspection module. Each scenario represents a realistic construction site for excavators. We compare the results of Trimble scanning Total Station ([11], a commercial land-surveying solution) and the proposed precise landscape inspection as shown in Table II. For comparison, we calculated the reprojection error onto the top-down view of the scene after aligning the coordinates of the two methods.

TABLE I: Computation time comparison between a commercial solution (scanning Total Station) and our method in sloped and flat ground scenarios such as Fig. 9 (a) and (d). The BA methods are capable of performing inspection within a larger FOV with less time consumption, and the proposed two-step algorithm further enhances computation speed.

Scenario	Method	Area [m^2]	Computation time [s]	Computation time per unit area [s/m^2]
	scanning Total Station	28	2542	90.8
Upward slope	Multi-view stereo — full BA	75	833	11.11
	Multi-view stereo — two-step BA (ours)	75	48	**0.64**
	scanning Total Station	75	3521	46.94
Flat ground	Multi-view stereo — full BA	101	978	9.68
	Multi-view stereo — two-step BA (ours)	101	59	**0.584**

Note: Bold underlined numbers represent the fastest among the three methods, in terms of computation time per unit area.

TABLE II: Quantitative comparison of 3D reconstruction accuracy and precision for the four validation scenarios shown in Fig. 9. The measurements obtained from Trimble scanning Total Station are used as the ground truth. The LiDAR-only method uses LiDAR pointclouds accumulated during the acquisition of multiple stereo images; the single-view stereo method reconstructs a 3D pointcloud with only one pair of images; and the multi-view stereo method is the proposed precise landscape inspection method.

		(a) Smooth slope	(b) Slope with a crater	(c) Uneven slope	(d) Flat ground
LiDAR only	Median	2.7	1.4	**0.8**	3.0
	StDev.	6.0	5.9	8.8	4.2
Single-view stereo	Median	1.5	1.5	2.8	9.1
	StDev.	13.4	4.2	**3.7**	4.0
Multi-view stereo (ours)	Median	**0.4**	**0.6**	**0.8**	**0.5**
	StDev.	**4.7**	**2.8**	7.0	**2.7**

Note: Unit is [cm]. Bold underlined numbers represent the best among the three methods. StDev. = Standard deviation.

$45°$ upward slope that is 3.5 m wide and 1.5 m high. Such an upward slope cutting task is one of the most commonly conducted excavation tasks in the first stage of construction. Our experiment site consists of up to 10 cm-sized gravel and soil.

The hardware is a 30-ton excavator, Hyundai HX300AL. For communication between onboard sensors and computer, Ethernet and CAN communication networks are employed. The excavator's bucket can scoop up to 1 m^3 soil in a single excavation. On the bucket tip are steel teeth of length 10 cm attached in order to facilitate earthcutting. The hydraulic actuators are controlled by spool valves. The positions of the spools determine the flow rate to each actuator. The spools are driven by the *pilot* hydraulic pressure, which is controlled by the electronic signal from the onboard computer through electro-hydraulic pressure reducing (EPPR) valves at 100 Hz. The 3D rotating LiDAR sensors used in the experiment are capable of measuring distance at 3-cm-level accuracy. The IMUs provide the posture of the arm part with accuracy of $0.1°$, and the rotary encoder measures the swing angle at $2°$ accuracy.

Since the amount of soil to be excavated exceeds the capacity of the bucket, the excavator needs to repeat dig cycles at different base positions. To attain the desired width

of the slope, we perform excavation at three different base locations by moving the crawler in the horizontal direction to the landscape.

At every base location, excavation is performed until the bucket tip reaches the tip grading threshold where we set the tolerance as 50 cm. Due to soil movement, tip grading sometimes results in an error in height more than the predefined tolerance. In that case, the excavator performs excavation and tip grading over again. Such a cycle repeats until the height error is uniformly smaller than the tolerance in the ROI.

B. Result and Discussion

Experimental results in Fig. 10 (a)-(b) show the excavation process including snapshots of estimated landscape (a1)-(a4), earthcutting result (b1), and the landscape inspection result (b2).

A qualitative investigation on the 35 m^2 region around the excavated landscape was carried out using the proposed inspection method, as shown in Fig. 10 (d). For precise landscape inspection, considering the shape of the ROI, we fixed the boom angle and adopted a swing-only motion of $60°$. The mean and standard deviation of the absolute value of the elevation error were calculated as 7.4 cm and 5.3 cm, respectively, where the elevation error was measured within
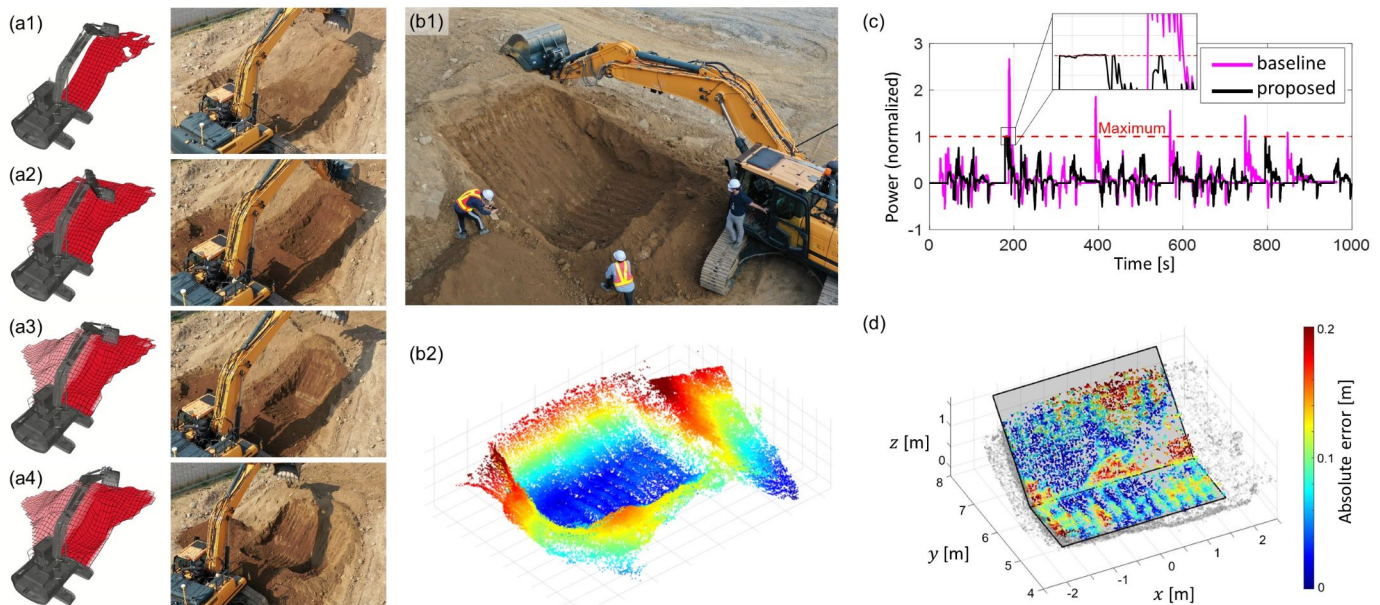
Fig. 10: Experimental result. (a1)-(a4) show landscape change as excavation proceeds. Excavation starts at the state in (a1), and the target landscape is established through an iterative digging cycle. (b1) Landscape after excavation terminates. (b2) Precise landscape inspection result after excavation. (c) Power constraint, one of the physical constraints, is shown to be satisfied through the proposed MPC throughout the execution time, while the constraint-free baseline algorithm fails to adhere to the constraint. (d) Comparison between the desired landscape and the result of precise landscape inspection for the 3.5 m × 3 m ROI. Excavation is accomplished within 10 cm accuracy.

the rectangular ROI of 3.5 m in width and 3 m in depth. The result shown in Fig. 10 (b2), (d) is sufficiently precise considering the excavator's size and the bucket type; it is at a similar level to that by professional operators.

Thanks to the MPC-based motion generation step which prevents violation of the physical constraints (Fig. 10 (c)), stable excavation can be accomplished without wobbling the bucket tip position. We found that, in the early stages of the earthcutting task, where there is a substantial amount of soil to be excavated and significant earth-bucket interaction forces are present, the bucket tip exhibits relatively large tracking error of up to 50 cm. However, as the operation progresses and the amount of soil to be excavated decreases, along with reduction in the load on the excavator, we observed that the excavator gradually achieves better precision in tracking the planned trajectory, resulting in the aforementioned centimeter-level earthcutting quality. The tip position error, however, largely relies on the performance of the low-level controller such as [22], [23]. This is out of the scope of this paper, but we believe excavation error can be further improved by enhancing the controller.

C. Computation Time Analysis

Finally, we conduct computation time analysis. All the proposed software are implemented and run on an onboard computer with a quad-core CPU at base clock 1.8 GHz and 8 GB RAM. First, the landscape estimation module takes measurements of the two 3D LiDARs running at 10 Hz and generates a 2.5D elevation map with a grid size of 40 cm × 40 cm in 5 Hz. Considering the rate of landscape change, this update rate is sufficiently fast to reflect changes in real-time. For precise landscape inspection, it takes 60 s for photo-shooting

(i.e., intentional swing motion), and the algorithm itself takes 25.8 s (or 0.74 s/m^2). Compared with the widely-applied Trimble scanning Total Station [11] which spends 1192 s (or 34.2 s/m^2), the proposed method consumes significantly less time while providing competent performance. A digging path is computed in 0.3 s in most cases, and the path-tracking MPC generating a local trajectory runs faster than 0.08 s. Since the computation time of the digging profile generation is faster than the float time between two sequences of soil dumping and returning to the excavation plane with swing motion, ceaseless excavation can be performed. Furthermore, the physical constraints can be satisfied, thanks to the fast running rate of MPC that enables real-time response to the change of cylinder pressure and the excavator configuration.

IX. CONCLUSION

A. Summary

This research proposed an integrated AES that combines real-time landscape estimation, excavation path generation, path tracking, and precise landscape inspection. In detail, we suggested a sensor arrangement to ensure sufficient FOV even during excavation regardless of the target landscape, a motion planning technique considering both geometric and physical constraints, and real-time landscape estimation and precise landscape inspection methods that rely on onboard sensors only.

These proposed techniques were validated on a real 30-ton hydraulic excavator. During experiments, the proposed methodology successfully executed an earthcutting task on a 45° upward slope, achieving accuracy within 10 cm through repetitive autonomous excavation tasks. The newly suggested post-excavation inspection technique completed its task within

one minute, showing an accuracy level of less than one centimeter. With the tens of times faster speed than commercial solutions in terms of computation time per inspection area, our method achieved a similar level of inspection accuracy.

B. Discussion and Future Work

While we successfully demonstrated completing a single excavation task autonomously using the proposed AES, there still remains the need for additional research for deployment of an AES in real-world construction sites.

In terms of perception, using cameras for landscape inspection poses drawbacks such as the inability to work during darkness and susceptibility to weather conditions. Improvement in this aspect is needed, including employing *backup* sensors. Regarding excavation motion generation, our method relies on the local planner to track the global planning generated when the digging cycle begins, but the excavator may get stuck in the presence of unperceived rigid undiggable obstacles in the soil, such as bedrock. Improvements to the local planner can be made to address this issue; for instance, creating and utilizing a 3D map of underground soil ease of excavation based on hydraulic pressure could be a viable solution.

To deploy AES in large-scale construction sites, the movement of the excavator base should be taken into account. This includes (but not limited to) the higher-level task planning on where to begin digging such as [24], cooperation with other excavators and heavy equipment such as dump trucks, and developing additional modules to distinguish drivable and non-drivable areas using onboard sensors.

REFERENCES

[1] "SafeAI: Autonomous mining & autonomous construction solutions," https://www.safeai.ai/, accessed: 2024-02-29.

[2] "Robots that build the world – Built Robotics," https://www.builtrobotics.com/, accessed: 2024-02-29.

[3] L. Zhang, J. Zhao, P. Long, L. Wang, L. Qian, F. Lu, X. Song, and D. Manocha, "An autonomous excavator system for material loading tasks," *Science Robotics*, vol. 6, no. 55, p. eabc3164, 2021.

[4] D. Jud, S. Kerscher, M. Wermelinger, E. Jelavic, P. Egli, P. Leemann, G. Hottiger, and M. Hutter, "Heap-the autonomous walking excavator," *Automation in Construction*, vol. 129, p. 103783, 2021.

[5] R. Mascaro, M. Wermelinger, M. Hutter, and M. Chli, "Towards automating construction tasks: Large-scale object mapping, segmentation, and manipulation," *Journal of Field Robotics*, vol. 38, no. 5, pp. 684–699, 2021.

[6] B. Cao, X. Liu, W. Chen, H. Li, and X. Wang, "Intelligentization of wheel loader shoveling system based on multi-source data acquisition," *Automation in Construction*, vol. 147, p. 104733, 2023.

[7] D. Jud, P. Leemann, S. Kerscher, and M. Hutter, "Autonomous free-form trenching using a walking excavator," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 3208–3215, 2019.

[8] R. L. Johns, M. Wermelinger, R. Mascaro, D. Jud, F. Gramazio, M. Kohler, M. Chli, and M. Hutter, "Autonomous dry stone: On-site planning and assembly of stone walls with a robotic excavator," *Construction Robotics*, vol. 4, no. 3-4, pp. 127–140, 2020.

[9] E. Jelavić, D. Jud, P. Egli, and M. Hutter, "Robotic precision harvesting: Mapping, localization, planning and control for a legged tree harvester," *Field Robotics*, vol. 2, pp. 1386–1431, 2022.

[10] D. Moon, S. Chung, S. Kwon, J. Seo, and J. Shin, "Comparison and utilization of point cloud generated from photogrammetry and laser scanning: 3D world model for smart heavy equipment planning," *Automation in Construction*, vol. 98, pp. 322–331, 2019.

[11] "Trimble Total Stations — Trimble Geospatial," https://geospatial.trimble.com/products-and-solutions/total-stations, accessed: 2024-02-29.

[12] Z. Yao, S. Zhao, X. Tan, W. Wei, and Y. Wang, "Real-time task-oriented continuous digging trajectory planning for excavator arms," *Automation in Construction*, vol. 152, p. 104916, 2023.

[13] Y. Yang, P. Long, X. Song, J. Pan, and L. Zhang, "Optimization-based framework for excavation trajectory generation," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1479–1486, 2021.

[14] E. Jelavic and M. Hutter, "Whole-body motion planning for walking excavators," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019, pp. 2292–2299.

[15] D. Lee, I. Jang, J. Byun, H. Seo, and H. J. Kim, "Real-time motion planning of a hydraulic excavator using trajectory optimization and model predictive control," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, pp. 2135–2142.

[16] Y. B. Kim, J. Ha, H. Kang, P. Y. Kim, J. Park, and F. C. Park, "Dynamically optimal trajectories for earthmoving excavators," *Automation in Construction*, vol. 35, pp. 568–578, 2013.

[17] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.

[18] J. Kim, C. Kim, Y. Han, and H. J. Kim, "Automated extrinsic calibration for 3D LiDARs with range offset correction using an arbitrary planar board," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 5082–5088.

[19] F. A. Bender, S. Göltz, T. Bräunl, and O. Sawodny, "Modeling and offset-free model predictive control of a hydraulic mini excavator," *IEEE Transactions on Automation Science and Engineering*, vol. 14, no. 4, pp. 1682–1694, 2017.

[20] B. Plancher, Z. Manchester, and S. Kuindersma, "Constrained unscented dynamic programming," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 5674–5680.

[21] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I 9*. Springer, 2006, pp. 430–443.

[22] J. Park, B. Lee, S. Kang, P. Y. Kim, and H. J. Kim, "Online learning control of hydraulic excavators based on echo-state networks," *IEEE Transactions on Automation Science and Engineering*, vol. 14, no. 1, pp. 249–259, 2017.

[23] S. Kim, J. Park, S. Kang, P. Y. Kim, and H. J. Kim, "A robust control approach for hydraulic excavators using μ-synthesis," *International Journal of Control, Automation and Systems*, vol. 16, pp. 1615–1628, 2018.

[24] J. Kim, D.-e. Lee, and J. Seo, "Task planning strategy and path similarity analysis for an autonomous excavator," *Automation in Construction*, vol. 112, p. 103108, 2020.