AFRL-AFOSR-JP-TR-2022-0039

Hybrid site sensing and human-multi-robot team collaboration for disaster relief at nuclear power plants

Cho, Yong
GEORGIA TECH RESEARCH CORPORATION
926 DALNEY ST NW
ATLANTA, GA, 30332
USA

05/31/2022
Final Technical Report

Air Force Research Laboratory
Air Force Office of Scientific Research
Asian Office of Aerospace Research and Development
Unit 45002, APO AP 96338-5002

REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE	**2. REPORT TYPE**	**3. DATES COVERED**	
20220531	Final	**START DATE** 20170421	**END DATE** 20201020

4. TITLE AND SUBTITLE
Hybrid site sensing and human-multi-robot team collaboration for disaster relief at nuclear power plants

5a. CONTRACT NUMBER	**5b. GRANT NUMBER**	**5c. PROGRAM ELEMENT NUMBER**
FA2386-17-1-4655		61102F
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**

6. AUTHOR(S)
Yong Cho

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
GEORGIA TECH RESEARCH CORPORATION
926 DALNEY ST NW
ATLANTA, GA 30332
USA

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**	**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**
AOARD UNIT 45002 APO AP 96338-5002	AFRL/AFOSR IOA	AFRL-AFOSR-JP-TR-2022-0039

12. DISTRIBUTION/AVAILABILITY STATEMENT
A Distribution Unlimited: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Robotic technology is important for creating spatial maps of harsh environments such as disaster sites. A virtual scanning environment is created by using a 3D CAD (Computer Aided Design) model of a nuclear power plant site and simulating various disaster scenarios such as reactor explosions and earthquakes using a physics engine. A virtual mobile robot is used to move around the disaster site and collect laser scans by carrying out raytracing on a voxel grid. This research also introduced an online segmentation method using Multi-View Context Pooling (MCP) for mobile robots to incrementally build a semantically rich 3D point cloud of the environment. During the post-processing stage for the acquired point cloud data, a three-step approach of segmentation, classification, and merging for object recognition is applied, such that each object is precisely represented as a cluster of points that have a unique class label

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**		
U	U	U	SAR	37

19a. NAME OF RESPONSIBLE PERSON	**19b. PHONE NUMBER** *(Include area code)*
JERMONT CHEN	315-227-7003

RESEARCH FINAL REPORT TO AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFOSR)		Date: 10/1/2020
	Report Period From: 4/21/2017	To: 10/1/2020
Project No. FA2386-17-1-4655	Project Title: **Hybrid site sensing and human-multi-robot team collaboration for disaster relief**	
Research Agency (s): Georgia Institute of Technology	Principal Investigator(s): Yong K. Cho, Associate professor, School of Civil and Environmental Engineering Jun Ueda, Associate professor, School of Mechanical Engineering	
Starting Date: 4/21/2017	Completion Date: 10/1/2020	

Summary

Robotic technology is important for creating spatial maps of harsh environments such as disaster sites. A virtual scanning environment is created by using a 3D CAD (Computer Aided Design) model of a nuclear power plant site and simulating various disaster scenarios such as reactor explosions and earthquakes using a physics engine. A virtual mobile robot is used to move around the disaster site and collect laser scans by carrying out raytracing on a voxel grid. This research also introduced an online segmentation method using Multi-View Context Pooling (MCP) for mobile robots to incrementally build a semantically rich 3D point cloud of the environment. During the post-processing stage for the acquired point cloud data, a three-step approach of segmentation, classification, and merging for object recognition is applied, such that each object is precisely represented as a cluster of points that have a unique class label. A deep learning architecture was designed to identify building elements such as columns, door, and stairs that may be potentially deformed in a disaster relief scenario. To simulate damaged objects encountered in post-disaster scenarios, synthetic deformations including bending and truncation were applied to a database of building element CAD models. This research approaches the task of detecting cracked regions in damaged structures using unsupervised learning techniques. Point-level features are first extracted from the point cloud data in the form of RGB color, intensity, normal vector, curvature, and neural network-based features. Then, K-means clustering is applied to identify outlier points that correspond to a cracked or damaged region. Robotic agents can also be remotely controlled to perform manipulation tasks for excavation, sample collection and repair work in disaster sites. It is difficult for human operators to efficiently and accurately carry out manipulation tasks through a teleoperation medium without clearly perceiving the pose of the robot and objects around it. Thus, this research proposes a context-aware 3D workspace modeling of remote equipment to improve perception and situational awareness for efficient teleoperation. The 3D workspace modeling process includes laser scanning, automated object detection, pose estimation, and 3D visualization to provide real-time feedback to the operator on the task progress. In terms of tele-operation

control, this research implemented control systems on unmanned excavator robots operated by the GHOST system at the Hanyang University. For low-level motion control, steps were taken to model the viscoelastic characteristics of the excavator system used as a remote manipulator. Accelerometers mounted on the cabin, crawler-base, and gripper/end-effector of the excavator fitted with the GHOST telemanipulation system show that for some motion tasks there is base-excitation of the excavator. For mid-level control, a controller was proposed that employs a form of oblique coordinate control. Finally, for high-level command generation, a method was formulated to segment the robot working area into multiple sub-regions based on the risk assessment and utilize quantifiable operator confidence/trust to actively merge manual operation commands and autonomous commands. To test the developed theory, a scale-model of an excavator is being outfit to be a robotic device. The system utilized a modular design, where the sensor systems and control theory-programming can be swapped between different mechanical and computer systems with minimal necessary changes. Human-interface is performed with a 3D mouse (3D Connexion, SpaceNavigator) to provide motion commands to the mobile excavation system. Further investigative study was conducted into the properties of the stability margins used to optimize the posture of the manipulator against the unknown base disturbances. An encrypted master-slave teleoperation system applying El-Gamal encryption was realized, and a data sharing scheme was provided for a two-key two-end-device (master and slave) system, and the computation of system inputs in ciphertext was performed.

Research Objective / Tasks:

Task 1-1: Hybrid 3D Laser Scan for Site Mapping
Task 1-2: Damaged/Deformed Object Recognition and Classification
Task 1-3: Context-aware 3D Workspace Visualization for Remote Operators
Task 3-1: Adaptive Gain Scheduling for Robust Tele-operation
Task 3-2: Shared Control for Scaled Hand-Arm Tele-operation

Status:

Task 1-1: 100%
Task 1-2: 100%
Task 1-3: 100%
Task 3-1: 100%
Task 3-2: 100%

Research Collaboration Efforts with the Korean Team:

Collaboration with The Hanyang University at Erica.

- Two graduate students from Dr. Cho's group, Mr. Pileun Kim, and Mr. Jingdao Chen, visited the Hanyang University Campus on May 21-30, 2018. The visit aims to test a mobile laser scanning and visualization system that will provide real-time point cloud representations of a robot workspace to help operators control grippers from a remote location.
    - The Georgia Tech team worked closely with Dr. Han's group at Hanyang University team throughout the trip to integrate the laser scanning and visualization system from Tech with the excavator system from Hanyang University.
    - From May 21-24, the research team worked on algorithm development and software deployment.
    - From May 27-30, the research team worked on data collection and real-time manipulation tests with the remotely-controlled excavator system.
    - A final presentation about the results and was made to the PIs at the conclusion of the trip

- Dr. Cho's group collaborated with Dr. Ahn's group at Hanyang University to perform disaster site risk analysis.
    - Dr. Ahn's group conducted structural simulations to create test data for damaged concrete structures
    - Dr. Cho's group used 3D point cloud data of the damaged structures to perform tilt estimation, displacement estimation, and crack detection
    - Dr. Ahn's group used the damage detection results to conduct risk analysis
- Two graduate students from Dr. Han's group, Mr. Dong Ik Sun, and Ms. Min Ji Kim, visited Georgia Tech Campus on February 5-8, 2018. Dr. Ueda and his graduate student, Michael Libretto met every day and discussed research progress and future plans during their visit. Several new research topics have been identified including classification of ground conditions, clawer oscillation, and excavator end-point compliance analysis and feedback control, to better interface Objective 3 (Ueda) with Objective 4 (Han)
- Dr. Yonghan Ahn (co-PI) at Hanyang Univ. visited Georgia Tech campus on March 21, 2018, to discuss the disaster site risk analysis based on the 3D map created by Dr. Cho's group. We decided to use a drone scanned disaster site map to quickly identify high-risk zones for the further detailed site investigation by a mobile robot.
- Dr. Ueda's group works with Dr. Han's group at the Hanyang University to develop algorithms to control remote excavators equipped with the GHOST teleoperation interface.
- Team members had online meetings regularly via Skype to collaborate on research papers.

Publications:

1. Zeng, S., Chen, J., and Cho Y. (2020). "User Exemplar-based Building Element Retrieval from Raw Point Clouds using Deep Point-level Features". Automation in Construction, Volume 114, June 2020, 103159
2. Park, J., Chen, J., and Cho Y. (2020). "Point Cloud Information Modeling (PCIM): an Innovative Framework for as-is Information Modeling of Construction Sites". ASCE Construction Research Congress (CRC) 2020, March 9-10, Tempe, AZ.
3. Chen, J., Kira, Z., and Cho. Y. (2020). "Learnable Region Growing for Class-Agnostic Point Cloud Segmentation". IEEE International Conference on Intelligent Robots and Systems (ICRA) , *under submission*
4. Chen, J., and Cho, Y. (2020). "Unsupervised Crack Segmentation from Disaster Site Point Clouds using Point Feature Clustering." Proceedings of Workshop of the European Group for Intelligent Computing in Engineering, EG-ICE
5. LiBretto, M., Ahn, Y. H., Han, C. S., Cho, Y. K., and Ueda, J. (2020). "Configuration Optimization for End-Point Stabilization of Redundant Manipulators With Base Flexibility." ASME. *Letters Dyn. Sys. Control*. April 2021; 1(2): 021001. https://doi.org/10.1115/1.4046684
6. Teranishi, K., Ueda, J., and Kogiso, K, (2020) "Encrypted Feedback Linearization and Motion Control for Manipulator with Somewhat Homomorphic Encryption," in IEEE AIM Conference, pp. 613-618, July 6, 2020
7. Teranishi, K., Ueda, J., and Kogiso, K, (2020), "Event-triggered approach to increasing sampling period of encrypted control systems," IFAC World Congress, 7/12-7/17
8. Chen, J. and Cho, Y. (2019). "Exemplar-based Building Element Retrieval from Point Clouds." International Conference on Smart Infrastructure and Construction (ICSIC), Churchill College, Cambridge, UK, July 8-9, DOI: 10.1680/icsic.64669.225
9. Kim, P., Price, L., and Cho, Y. (2019). "UAV-UGV Cooperative 3D Environmental Mapping." Proceedings of the ASCE 2019 International Conference on Computing in Civil Engineering (i3CE), Atlanta, GA, USA, June 17-19, DOI: doi/pdf/10.1061/9780784482438.049

10. Chen, J., and Cho, Y. (2019). “Detection of Damaged Infrastructure on Disaster Sites using Mobile Robots.” Proceedings of the 2019 IEEE 16th International Conference on Ubiquitous Robots (UR), Jeju, Korea, June 24-27.
11. Chen, J., Kira, Z., and Cho, Y. (2019) “Deep Learning Approach to Point Cloud Scene Understanding for Automated Scan to 3D Reconstruction.” ASCE Journal of Computing in Civil Engineering, in press.
12. J. Chen, Y. K. Cho and Z. Kira, "Multi-View Incremental Segmentation of 3-D Point Clouds for Mobile Robots," in IEEE Robotics and Automation Letters, vol. 4, no. 2, pp. 1240-1246, April 2019.
13. Chen, J., Kim, K., Cho, Y., Lee, J. S., Kim, B., Ahn, Y. H., Kang, J. “Nuclear Power Plant Disaster Site Simulation using Rigid Body Physics.” 2019 ASCE International Conference on Computing in Civil Engineering.
14. Teranishi, K., Kusaka, M. Shimada, N. Ueda, J. and Kogiso, K. (2019) “Secure observer-based motion control based on controller encryption,” American Control Conference, pp. 2978-2983
15. Qiu. Y. and Ueda, J., (2019) “Encrypted Motion Control of a Teleoperation System with Security-Enhanced Controller by Deception,” in ASME 2019 Dynamic Systems and Control Conference, In Dynamic Systems and Control Conference, vol. 59148, p. V001T07A006.
16. Park, J.S., Kim, P., Cho, Y., and Fang, Y. (2018)." Automated Collaboration Framework of UAV and UGV for 3D Visualization of Construction Sites." Proceedings of the 18th International Conference on Construction Applications of Virtual Reality (CONVR2018), Auckland, New Zealand, Nov 22-23.
17. Kim, P., Chen, J., and Cho, Y. (2018). "SLAM-driven robotic mapping and registration of 3D point clouds." Automation in Construction
18. Kim, P., Chen, J., Cho, Y. (2018). "Autonomous Mobile Robot Localization and Mapping for Unknown Construction Environments." ASCE Construction Research Congress (CRC) 2018, pp.147-156, April 2-4, New Orleans, LA
19. Chen, J., Cho, Y., and Ueda, J. (2018). "Sampled-Point Network for Classification of Deformed Building Element Point Clouds." Proceedings of the 2018 IEEE Conference on Robotics and Automation (ICRA), 21-25 May 20018, Brisbane, Australia.
20. Chen, J., Kim, P., Cho, Y., and Ueda, J. (2018). “Object-sensitive potential fields for mobile robot navigation and mapping in indoor environments.” Proceedings of the 15th International Conference on Ubiquitous Robots (UR), June 26-30, 2018, Honolulu, Hawaii, USA.

Detailed Research Progress:

Task 1-1: Hybrid 3D Laser Scan for Site Mapping

Disaster Site Scanning Simulation

This research proposes a method for collecting laser scan data from disaster sites using mobile robots and detecting damaged infrastructure from the acquired point clouds. Due to the difficulty of obtaining data from real-world disaster sites, a virtual disaster environment is used in this study. The virtual environment is obtained by creating a 3D CAD (Computer Aided Design) model of a nuclear power plant site and simulating various disaster scenarios such as reactor explosions and earthquakes using a physics engine. A virtual mobile robot is used to move around the disaster site and collect laser scans by carrying out raytracing on a voxel grid.

Two types of disaster scenarios are considered in this study: (i) reactor explosion and (ii) earthquakes. A 3D environment is created containing the nuclear power plant site. Next, the Bullet [1] physics engine is

used to simulate rigid body physics in the 3D environment. The technique of Voronoi cell fragmentation is used to decompose the 3D model into smaller rigid bodies. A series of forces are then applied to the rigid bodies as a result of the disaster scenario. Each post-disaster 3D model is then converted into a point cloud format by randomly sampling points along each surface of the mesh model. The point cloud is further stored in a voxel grid with a resolution of 0.5m to speed up data access.

A virtual mobile robot is used to move around the disaster site and collect laser scans. The mobile robot is placed on the ground plane with the laser scanner 1m above ground. Using the voxel grid from the previous section, a ray-tracing routine is carried out to compute the result of each laser scan. Whenever a ray intersects an occupied voxel, the corresponding 3D point is added to the scanned point cloud. The laser scans are thus continuously accumulated to build a 3D point cloud of the scanned environment. Note that the robot is assumed to have Simultaneous Localization and Mapping capability so that it can estimate its location at each point in time during the scanning process and accurately build a map of the environment.

Damage detection on a disaster site is performed by comparing the 3D geometry of post-disaster site conditions to that of the original site. This process assumes that a 3D representation of the original site is available, whether from 3D CAD designs or pre-disaster laser scans. As the mobile robot is scanning the environment and collecting point cloud data, the original point cloud from the pre-disaster site and the laser-scanned point cloud from the post-disaster site are overlaid on the same coordinate system. The point clouds are then voxelized to a grid of 0.5m resolution so that points have even spacing in the point cloud. Each point in the overlaid point cloud is then classified into one of three categories according to the following rules:

- Background: Points that are present in both the pre-disaster site and post-disaster site.
- Debris: Points that are present in the post-disaster site but not present in the pre-disaster site.
- Damage: Points that are present in the pre-disaster site but not present in the post-disaster site.

To eliminate noise in the classification results, a smoothing process is further applied so that the classification of each point is consistent with that of its neighboring points. Figure 1 shows an example of the damage detection results on the 3D point cloud dynamically scanned by a mobile robot. As the mobile robot moves around the disaster site and acquires point cloud data, the points are automatically labeled according to whether they match the pre-disaster site geometry. At the end of the scanning process, points that are present in the pre-disaster site but not detected in the post-disaster laser scans are labeled as "damage." Note that the damage detection results may be incomplete since there may be inaccessible regions in the disaster site as well as errors caused by occlusion and clutter. However, the qualitative damage detection results can still be visually presented to the disaster relief team so that critical tasks such as risk assessment and formulating search and rescue plans can be performed.

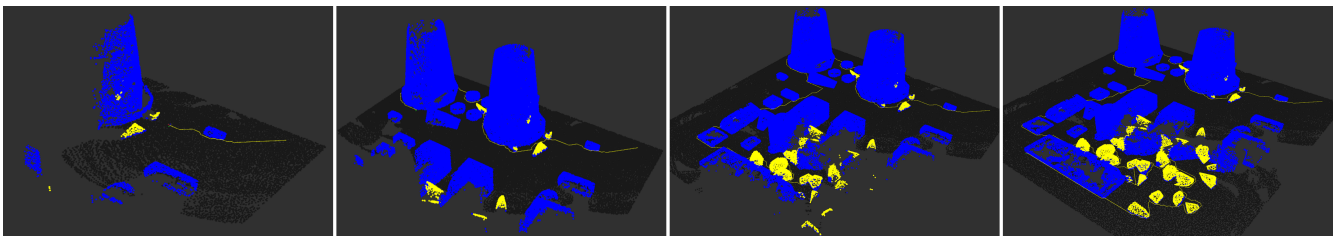


Figure 1: 3D point cloud dynamically scanned by a mobile robot in a simulated nuclear power plant site. Yellow points indicate debris whereas blue points indicate undamaged structures.

Multi-view Incremental Segmentation

Mobile robots need to create high-definition 3D maps of the environment for disaster relief. Accurate semantic processing of the acquired 3D point cloud is critical for allowing the robot to obtain a high-level understanding of the surrounding objects and perform context-aware decision making. Existing techniques for point cloud semantic segmentation are mostly applied on a single-frame or offline basis, with no way to integrate the segmentation results over time. This research proposes an online method for mobile robots to incrementally build a semantically rich 3D point cloud of the environment.

The instance segmentation process assigns a class label and instance label to each point in the scanned point cloud. The class label is computed directly by a neural network whereas the instance label is determined by agglomerative clustering. The segmentation results are progressively updated in a dynamic fashion as the robot traverses its environment and acquires laser scan data, allowing real-time object recognition and improved scene understanding for the robot. The proposed work uses a novel multi-view context pooling (MCP) module to aggregate information from previous views before performing instance segmentation. The key idea of this work is to use contextual information from not all, but a select few of the previous scans, to improve segmentation of the current scan. This online incremental segmentation framework allows new scans to be processed within tenths of a second compared to half a minute if segmentation of the entire point cloud were to be recomputed with an offline method.

The proposed incremental segmentation method involves a neural network architecture, named Multi-view Context Pooling Network (MCPNet), that jointly predicts an instance label and a class label for each scanned point. The class label describes the semantic class of a point (i.e., wall, door, table, etc.) whereas the instance label is a unique ID such that only points from the same object instance have the same ID. In order to perform instance segmentation in an online setting, it is important to keep track of newly scanned points and their relationship to previously scanned points in a global coordinate system. Thus, a voxel grid with a grid size of 0.1m is used as a lookup-table to store the point coordinates and their corresponding segmentation results. The voxel grid ensures that points are only added to the lookup-table when new regions are scanned and prevents the data from growing too quickly since the point cloud is limited to one point per voxel. The voxel grid is also key to efficiently retrieving neighbor points and context points for segmentation.

Each input scan acquired by the robot is first pre-processed to retain points only in an area of radius 2m around the robot. This is due to the fact that scan regions that are too far away from the robot are too sparse to make accurate predictions about the object class. Thus, faraway points are discarded for the current scan but can still be included in a future scan as the robot moves to a closer location. Next, the point coordinates are normalized so that the x-y coordinates are centered around the robot and the z-coordinate is zero at the floor level. The points are then divided into batches of Nx6 matrices, where N is the batch size and the columns are X-Y-Z-R-G-B values. If there are more than N points in the current scan, data will be passed to the network in multiple batches, where the last batch is sampled with replacement.

Figure 2 shows an outline of the proposed MCPNet architecture used to process the resulting input matrices. The input matrix is first passed through a 1D convolution layer to form an intermediate Nx200 feature matrix. The network then splits into two branches: the lower branch for classification and the upper branch for clustering, which will be described in detail in the next subsection. The lower branch uses a max pooling function to compute a global feature vector representing the input points. This is then concatenated with the feature embedding from the upper branch and passed through another 1D convolution layer to compute class probabilities for each point, represented as an Nx13 matrix since there are 13 classes.

The proposed architecture also incorporates a Multi-view Context Pooling (MCP) module which incorporates contextual information from previous views to points in the current scan to improve the classification accuracy. The input to the MCP module is an NxMx6 tensor that stores context points for

each of the N points in the current input. A context point for point p_i is defined as any point from previous views that is at most three cells away from p_i in the global voxel grid. The three-cell distance threshold is set so that only points from previous views that are in close proximity to the current view are used as context. M points are thus randomly sampled from among all the context points accumulated from past scans. The input tensor is passed through two layers of 1D convolutions as well as a max pooling layer. An Nx200 matrix forms the output of the MCP module and is concatenated to the Nx6 input matrix of the main network.

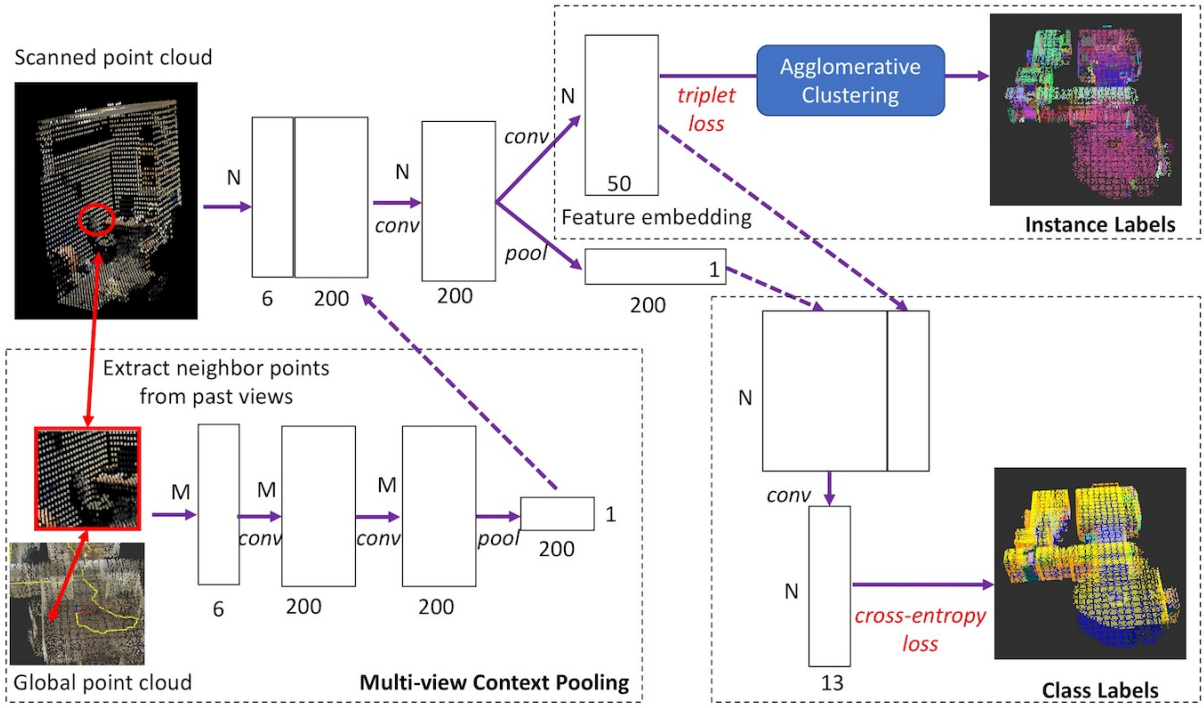


Figure 2: MCPNet architecture for incremental instance segmentation of 3D point clouds. The network consists of two branches: class labeling and instance labelling

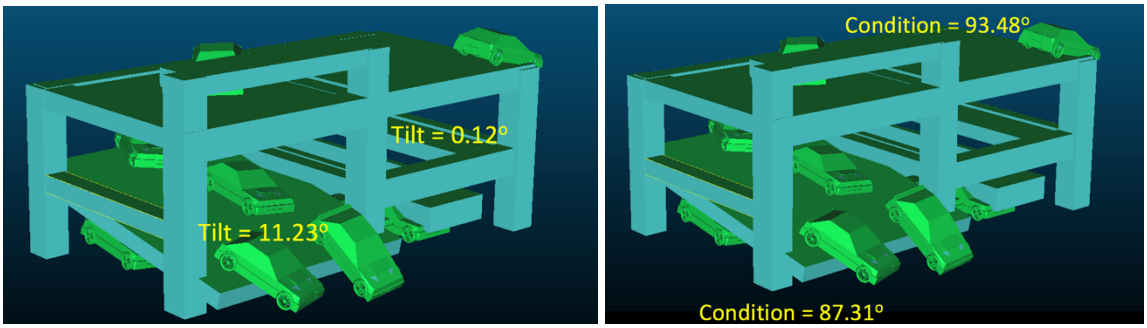
Field Test at Guardian Centers

To verify the proposed method for robotic scanning at disaster sites, the Guardian Centers disaster training facility was selected that has simulated damaged structures from earthquakes, hurricanes, or terrorist attacks to carry out experiments of the proposed method. This environment allows investigating the ability of the inspection robots to complete time-critical scanning missions. In environments such as disaster sites after an earthquake or explosion, spatiotemporal data is challenging to acquire since these environments are cluttered and difficult to navigate in. To address these problems, this research proposed a robotic navigation and registration framework that combines dynamic and static scan processes. During the dynamic scan, the mobile robot detects obstacles and computes efficient scan positions. Based on the results of the dynamic scan position estimation, high-resolution static scanning is performed under stationary conditions. The advantage of using the proposed framework over the prior works are as follows: (1) it automates the point cloud data acquisition process, (2) it works even with low overlapped areas between consecutive scans and without using targets or features in a point cloud; and (3) it requires less number of scans to reconstruct large sites, thus reducing time and cost for data collection.

Figure 3a shows our proposed ground robot collecting laser scan data next to a damaged concrete structure at Guardian Centers. Figure 3b shows the combined 3D point cloud of the damaged structure after registering the individual laser scans. From this site, a parking deck consisting of several wrecked cars and a tilted slab was selected for further analysis. As shown, in Figure 4, a 3D CAD model of the parking deck was constructed and tilt estimation was carried out by measuring the angle deviation of each building component whereas damage condition estimation was carried out by comparing the volume deviation of building components.

(a) (b)

Figure 3: (a) Ground robot scanning a damaged concrete structure at Guardian Centers (b) acquired 3D point cloud of the damaged structure



(a) (b)

Figure 4: (a) Tilt estimation and (b) damage estimation for a damaged parking deck structure at Guardian Centers

Task 1-2: Damaged/Deformed Object Recognition and Classification

Deep Learning Approach to Point Cloud Scene Understanding

Current methods for 3D object recognition are problematic due to the choice of object representation: (i) if classification is performed at the point level (i.e., semantic segmentation), there is no sense of object boundary that enables an observer to clearly define the separate entities in a point cloud scene, (ii) if objects are represented by 3D bounding boxes, accuracy issues may arise in situations with non-rectangular objects or when bounding boxes overlap, since bounding boxes only provide a coarse approximation of object boundaries. In addition, most works have also been benchmarked on synthetic data without significant testing on real-world 3D point cloud data, which is compounded by problems such as noisy data, low resolution, occlusion, clutter, and registration errors. This research proposes a three-step approach of segmentation, classification, and merging for object recognition from indoor point cloud scenes, such that

each object is precisely represented as a cluster of points that have a unique class label. The proposed method is evaluated quantitatively using the S3DIS dataset [2] and qualitatively with laser-scanned point cloud data of construction sites.

To perform segmentation, the point cloud is represented as a graph structure, with vertices representing 3D points and edges representing proximity between points. In this study, the point cloud is first subsampled with a voxel grid of 0.1m to normalize the point density and edges are formed between points that are within 0.2m Euclidean distance apart. In general, increasing the voxel size and increasing the distance threshold leads to under-segmentation whereas decreasing the voxel size and decreasing the distance threshold leads to over-segmentation. However, the values used in this study were empirically found to result in the best performance. Next, binary classification is performed based on edge features such that only edges between points on the same object are retained. The edge features are computed as follows, based on the XYZ-coordinates and RGB-color channels of both points connected by the edge:

- Minimum, maximum, and mean of the Z coordinate
- Minimum, maximum, and mean of the R, G, B channels
- Absolute difference in X, Y, Z coordinates
- Absolute difference in R, G, B channels

The intuition behind this formulation of edge features is that the points from the same object should have similar spatial coordinates and RGB color values, whereas the points from different objects should have larger differences in spatial coordinates and RGB color values. Using these edge features, a deep learning framework can then be applied to assess the likelihood that a point pair originated from the same object.

Once object segments are extracted from the point cloud scene, a modified version of PointNet [3], a deep neural network architecture for point clouds, is used to predict the class label of each object segment. The PointNet architecture is used because it allows for classification of a set of unstructured points that is invariant to the order of points. More specifically, it uses a pooling operation (i.e., max pooling) that is order-invariant to aggregate local features from each point to obtain global features.
PointNet receives as input a cluster of points and outputs a category prediction among 13 classes (clutter, beam, board, bookcase, ceiling, chair, column, door, floor, sofa, table, wall, window). However, the basic implementation does not perform well because some object segments may have as few as 30 points due to over-segmentation in the previous stage. The low number of points is insufficient to accurately predict the type of object, even for a human observer. To overcome this issue, the input point cloud is concatenated with surrounding points by extending the bounding box of the object by 1m in the X, Y and Z directions in order to provide context. For example, a small horizontal planar patch of points may be ambiguously classified as either a ceiling instance, a table instance, or a floor instance. However, incorporating contextual information from surrounding points enables the classifier to learn patterns in the point cloud geometry that indicate particular classes of objects; for example, ceiling points are located on top of walls and floors, whereas table points are located in between ceiling points and floor points.

After the second stage of the pipeline, the detected objects in the point cloud scene are still fragmented due to over-segmentation in the first stage of the pipeline. Even though each individual object segment may be classified correctly at the point level, they are still not represented correctly at the object level (e.g., a single wall may be fragmented into two or more walls segments). Thus, in this stage of the pipeline, object segments that are potentially from the same object are merged together to create unified objects. The following criteria are used to determine whether two object segments should be merged together:

- Both objects must have the same class
- The bounding box of both objects should be within 0.05m of each other.
- In the case of walls, the normal vectors of the two objects must be close to parallel

The final results of object segmentation, classification, and merging from point clouds are visualized for an indoor scene in Figure 5. The subfigures shown are from top to bottom, left to right are: (i) original point cloud; (ii) segmentation results (each segment is shown in a different color); (iii) classification results (yellow: ceiling, dark blue: floor, orange: wall, green: door, red: bookcase, pink: beam, light blue: column, gray: clutter); and (iv) object merging visualized with bounding boxes.

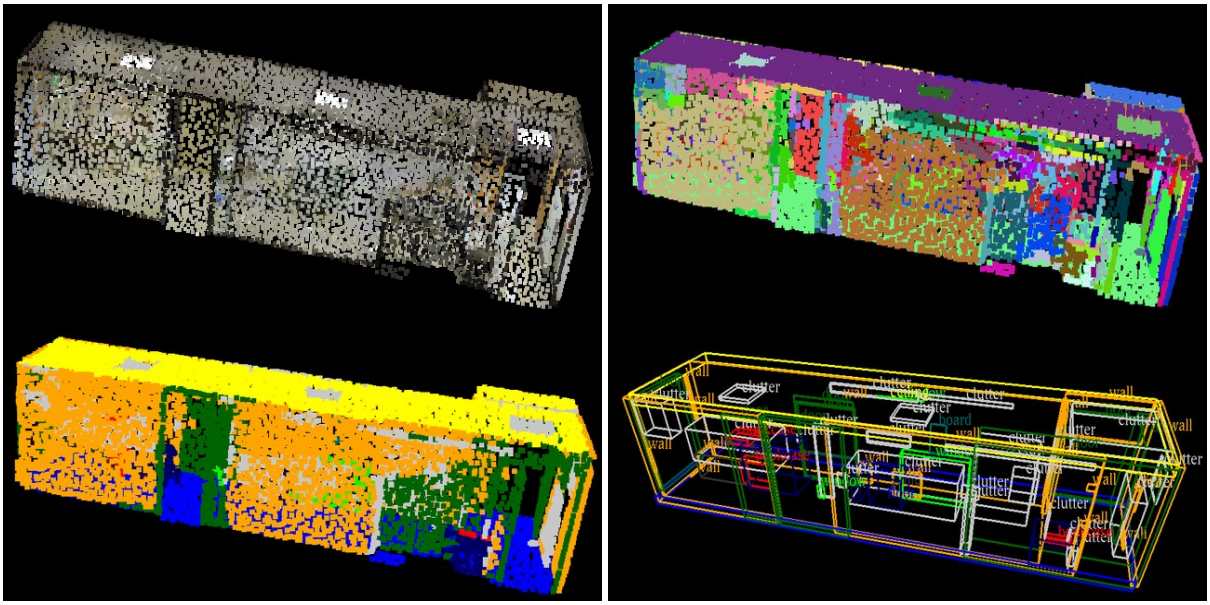


Figure 5: Visualization of segmentation and classification results on a hallway in Area 2 of the test dataset (top: original point cloud and segmentation results, bottom: classification results and merging results)

Classification of Damaged Building Elements

This study focuses on identifying building elements such as columns, door, and stairs that may be potentially deformed in a disaster relief scenario. The proposed method uses point clouds obtained from a database of building element mesh models to train a 3D object classification network. The classification network uses several regularization techniques such as point sorting, resampling, and orientation normalization to account for various distortions of 3D object point clouds. The accuracy of the classifier was evaluated by using original test sets as well as test sets derived from synthetically deformed objects. Finally, the classifier was qualitatively evaluated on point clouds of a reinforced concrete structure used for seismic testing.

First, this research aims to exploits the wealth of 3D data available in online repositories to learn an accurate 3D object classifier for building elements. A new dataset to synthesize point clouds from publicly available building element models (BIMNet10). This dataset is publicly available at http://github.com/jingdao/BIMNet10. BIMNet10 is a custom-built dataset consisting of mesh models of common building elements such as beams, columns, doors, fences, and stairs, obtained from Google's 3D Warehouse [7]. BIMNet10 contains 10 object classes, with 153 training samples and 47 test samples.

The input mesh models were converted into point clouds by randomly sampling points from each face of the model at a density of 2% of the model size. Next, the point clouds were normalized such that the centroid is located at the origin and the maximum radius is set to one. The point clouds were then uniformly down-sampled to match the number of points required as input to the classification network.

In order to simulate the deformations encountered for objects in post-disaster scenarios, this study applies synthetic deformations to the test set of BIMNet10. Figures 6 illustrates the different types of synthesized operations applied to the point clouds to build the deformed object datasets. The results of individual deformations (center 4 images) as well as that of combined deformations (rightmost image) are shown in Figure 6. The following explains the method of applying each type of deformation:

1) **Rigid Transformation:** Rigid transformation consists of translation, rotation, and reflection operations which maintain the shape and size of an object. In this study, only rotation operations are considered since (i) most object classifiers already re-center the object to account for translations and (ii) most real-world objects exhibit reflectional symmetry. Synthesized rotations are applied to the point clouds about the z-axis with the angle randomly selected between 0 and 2π.
2) **Outlier Noise:** Point cloud noise may originate from sensor measurement spikes, data registration errors when combining multiple scans, as well as segmentation errors. This is synthesized by appending several outlier points randomly placed in a cube surrounding the original object. The number of outlier points is 10% that of the original point cloud.
3) **Bending:** Bending occurs when excessive force is applied to a building element made from malleable materials such as metals. Bending displacements are generated as a curve in the form of $\delta x = A \cosh(s\, x)$ where x consists of the point coordinates normalized from -1 to 1 and s is a randomly-selected scale factor that controls the degree of bending. The displacements are applied along the longest axis of the object (e.g., the x-axis for a beam and the z-axis for a column)
4) **Truncation:** Truncation is normally encountered when an object breaks apart as a result of collisions or shocks. In this study, objects are truncated by randomly setting cutoff limits between 75% and 100% of the x, y, and z-axes and filtering out points that lie outside of those limits.

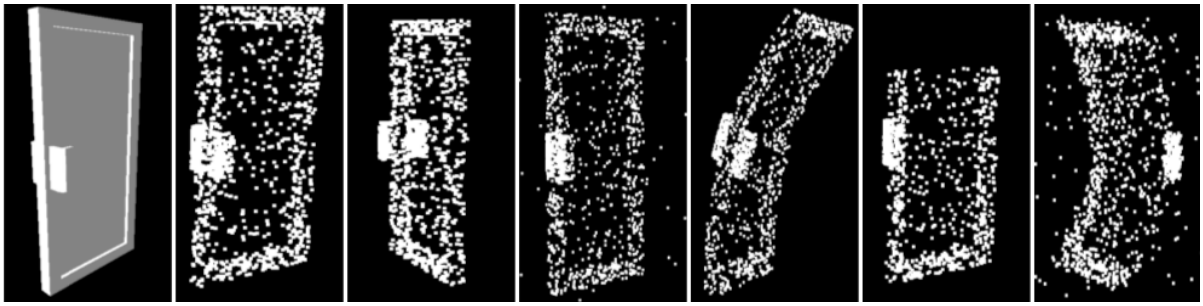


Figure 6: Synthesized deformation for a door object. From left to right: (i) mesh model (ii) point cloud (iii) rotated (iv) distorted with noise (v) bent (vi) truncated (vii) combined deformation

The architecture design used in this study is shown in Figure 7 below. The input point cloud is first converted into a two-dimensional array which represents a Sampled-Density Grid (SDG). The Sampled-Density Grid is constructed by sampling a random set of N points and taking the normalized x, y, z coordinates. This input is then passed through a series of 5 convolutional and pooling layers which successively reduce the spatial resolution. The convolutional filters have spatial dimensions of 1x3 for the first layer and 1x1 for the subsequent layers and are applied with a stride of 1. Each convolutional layer is followed by a ReLU layer and an average pooling layer with width 3 and stride 3. Thus, the number of input points must be at least $3^5 = 243$. After the last pooling layer, the input tensor is reshaped into a one-dimensional feature vector. This is then passed to a 10 or 40-length fully-connected prediction layer, depending on the number of object classes. The proposed deep learning architecture was implemented on

Tensorflow and trained using the ADAM optimizer with a batch size of 128. The learning rate was set to 0.01 and halved every 1000 batches.

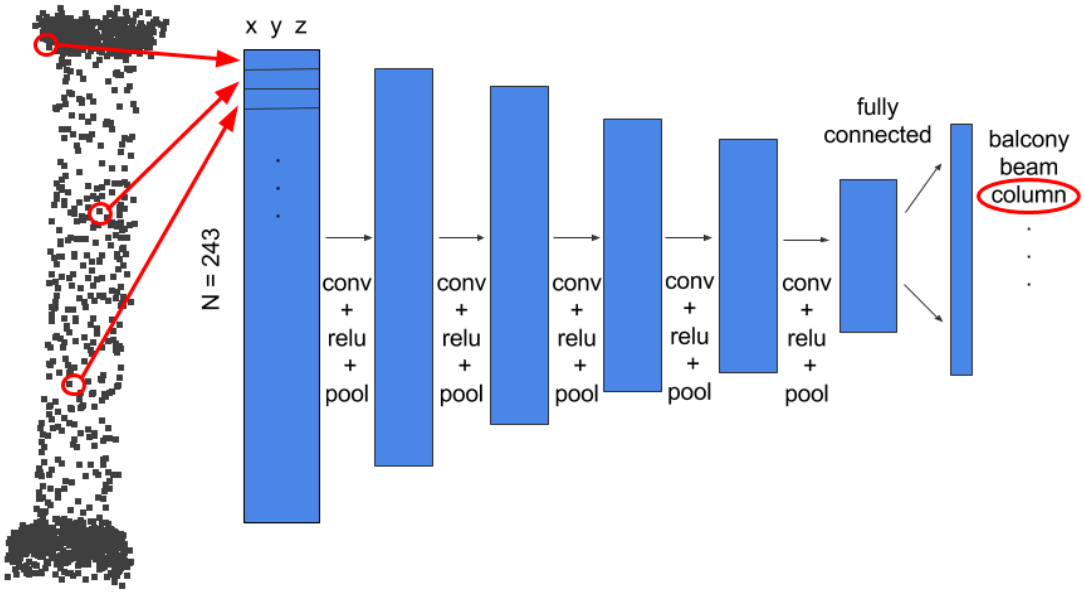


Figure 7: Network architecture for building element classification

The method was evaluated using point cloud data from a two-story reinforced-concrete building used for earthquake testing (Figure 8). The building point cloud was segmented according to the following procedure: (i) plane segmentation was applied to detect horizontal planes (ii) Euclidean clustering was used form clusters based on point-to-point distances. The proposed classification network was applied to assign a class label to each segmented object. In Figure 9, a classification result is shown if the class score is greater than 0.5, otherwise, the object is considered as “background”.

Figure 8: Original point cloud of seismic testing structure

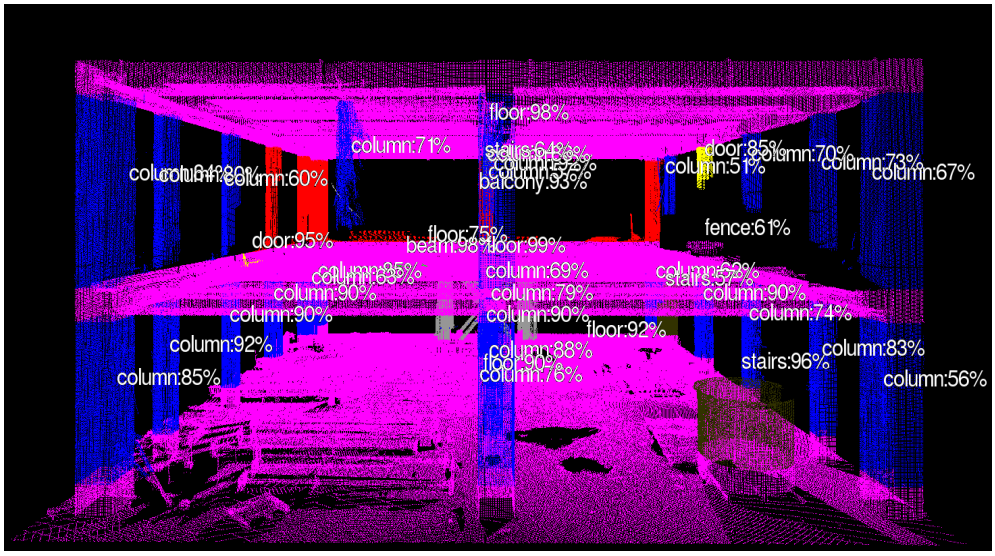


Figure 9: Building element point cloud classification results

Unsupervised Learning for Crack Detection

LiDAR technology plays an important role in disaster relief operations in scanning affected infrastructure and identifying the hazard level of different buildings on the disaster site. LiDAR technology is able to capture the 3D geometry of a building at mm-level accuracy, thus various deformations such as cracking and spalling on a building structure can be detected and analyzed from point cloud data. Once damaged regions have been identified, infrastructure damage information can then be compiled and categorized according to typical damage modes and compiled into an as-damaged Building Information Model. These models contain detailed and semantically rich information about the location and shape of infrastructure damage present in a building that can be useful to civil engineers and reconstruction and recovery (R&R) personnel. In order to generate these models, it is important to have automated methods that can detect damaged structures accurately and efficiently.

In contrast to LiDAR-based damage modeling, conventional methods for structural damage analysis rely on a long sequence of steps including field observations and vibration analysis that is very labor-intensive. In addition, these methods depend on having physical access to the disaster site which may not be practical for hazardous areas that can only be accessed remotely (e.g. by mobile scanning robots). Other methods can perform damage detection from live sensor data such as RGB images. However, these methods rely on having a large amount of labeled training data which is difficult to obtain when there are many different types of structural damages.

To achieve the goal of structural damage detection that can be used to reconstruct the 3D as-damaged geometry of a disaster site, this research proposes an unsupervised learning method to perform crack segmentation directly from LiDAR-scanned point cloud data without explicit training on the task of crack segmentation. The advantage of an unsupervised detection method is that it does not require training data of annotated cracks but can still adapt to different distributions of test data. Point-level features are first extracted from the point cloud data in the form of RGB color, intensity, normal vector, curvature, and neural network-based features. Then, K-means clustering is applied to identify outlier points that correspond to a cracked or damaged region. The proposed method is evaluated based on laser-scanned point clouds from the 2015 Nepal earthquake.

This study implements the K-means clustering algorithm due to its speed and interpretability. K-means clustering works by dividing the input data into K number of clusters such that each cluster contains data points which have roughly similar features. The intuition behind this step is that points with regular features will be grouped into large clusters whereas points with irregular features will be grouped into smaller clusters or appear as outliers. To perform crack segmentation, K-means clustering is applied to the features computed for each point and the resulting cluster with the fewest number of points is assigned as the cracked or damaged region. In this study, the K-parameter is set to 2 for the curvature-based method, 5 for the neural network-based method, and 3 for all other methods (depending on which setting results in the best clustering).

Figure 10 below show a qualitative comparison between the crack segmentation results for a concrete column. The qualitative results show that the RGB-based feature (third column) and neural network-based feature (rightmost column) show the most similar crack segmentation results to that of the ground truth (second column). Whereas, the results using other features such as intensity, normal vector, or curvature are more susceptible to noise and contain many instances of false positives and false negatives.

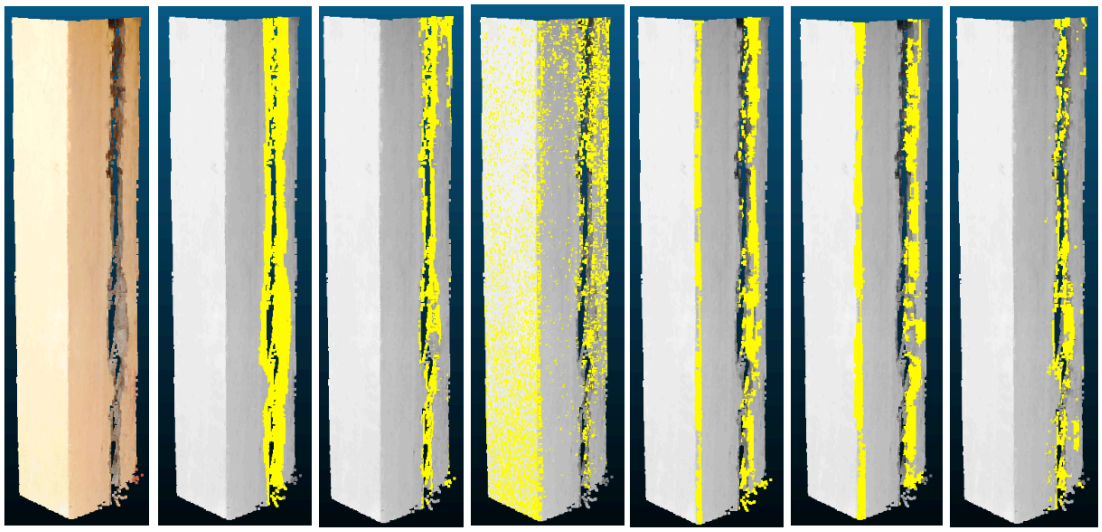


Figure 10: Visualization of crack segmentation results with different point features (yellow points indicate crack location). The subfigures are: (i) original point cloud, (ii) ground truth, results from (iii) RGB color (iv) intensity (v) normal vector (vi) curvature (vii) neural network

Task 1-3: Context-aware 3D Workspace Visualization for Remote Operators

Excavator Workspace Modeling and Visualization

Robotic agents have enormous potential to be used to perform manipulation tasks for excavation, sample collection and repair work in remote areas. In hazardous environments such as nuclear power plants or post-earthquake disaster sites, it is common for these robots to be teleoperated by human operators from a remote location. Such challenging conditions require a high level of situational awareness from the operator. It is difficult for human operators to efficiently and accurately carry out manipulation tasks through a teleoperation medium without clearly perceiving the pose of the robot and objects around it.

Research and field studies at major disaster relief operations such as the World Trade Center collapse showed that when mobile robots were deployed in confined spaces, the lack of perceptive data processing capability reduced the robotic skill set and added to the operator's cognitive responsibilities. Armed with only a raw video feed with noisy and blurry images, operators and rescuers were unable to keep track of where the robots searched and the conditions during the deployment. This problem of deficient perceptive information is amplified in the case of grasping tasks. Conventional teleoperation systems make use of only a video camera that provides 2D images to the operator and the lack of depth perception makes it difficult for the operator to estimate the size and distances of unknown objects. Moreover, visual cameras do not work well at night or in adverse weather conditions. As a result, operators require a significant amount of trial and error to correctly control the robot to complete a grasping task.

An integral part of intelligent perception is transforming sensor data into knowledge and expressing that knowledge as information for use by other members in a human-robot team. The idea of workspace modeling is to create a 3D representation of the robot and its surrounding environment containing both semantic and geometric information that can be shared among all human and robotic agents in the operational team. The workspace model is constructed in real-time by processing the raw sensor data, organizing it, and labelling relevant objects. Then the workspace model is used to provide visual feedback to the operator on the task progress.

This research proposes a context-aware 3D workspace modeling of remote equipment to improve perception and situational awareness for efficient teleoperation (Figure 11). In contrast to contemporary methods which only provide ego-centric views of the environment to the remote operator, this research proposes an observer-centric view with automated state estimation and visualization of the relative pose between the end effector and salient objects. In addition, this research was validated at a large-scale outdoor environment with real construction equipment, compared to existing methods which were more focused on small to medium scale indoor environments. To summarize, the contributions of the proposed work are as follows: (i) designed a 3D workspace modeling and visualization scheme for teleoperated construction equipment (ii) developed an accurate algorithm for pose estimation of articulated equipment without prior dynamic models (iii) demonstrated the effectiveness of the 3D workspace modeling system with field trials using a multi-excavator system teleoperated in real-time.

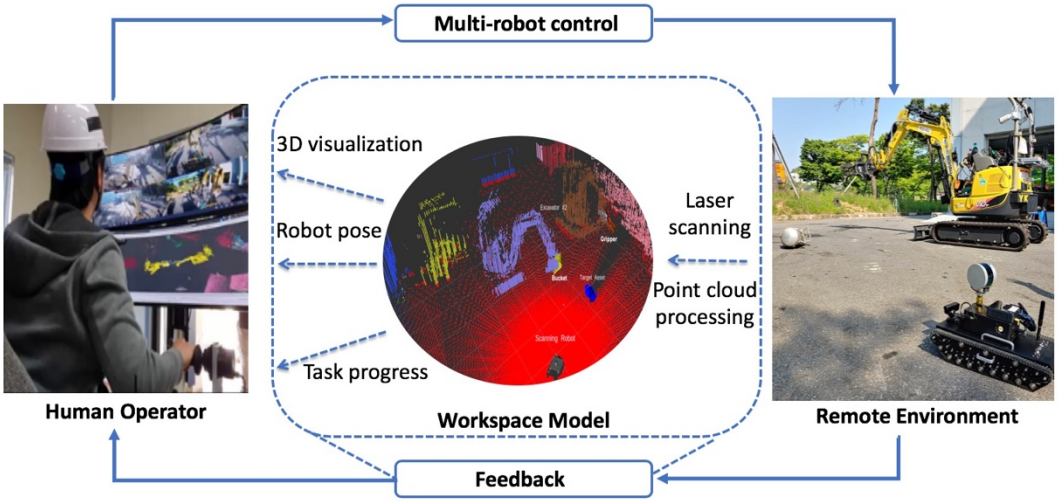


Figure 11: Feedback loop for teleoperation with 3D workspace modeling

The right image in Figure 11 shows the remote laser scanning system used to acquire a 3D reconstructed model of the excavator workspace. The Velodyne VLP-16 LiDAR device, commonly used for autonomous vehicles and advanced driver assistance systems, was mounted on a small teleoperated tracked mobile robot

and utilized to acquire real-time laser scans of the site. The VLP-16 has a range of up to 100 meters with 15º vertical field of view. This means that the vertical angular resolution is only 2º because the VLP-16 has only 16 scan lines (channels) in the vertical direction. Thus, the scanned point cloud will be overly sparse for regions far away from the scan origin. For this reason, this study implemented an improved scanning system by installing the VLP-16 at 90º sideways and spinning it continuously with a stepper motor. In this way, the limited vertical resolution (now horizontal resolution) can be mitigated by rotating the scans horizontally, thus creating a higher resolution point cloud.

The point cloud scene is segmented into smaller units corresponding to individual objects to enhance visualization of the scene. To meet the real-time constraint, the fast segmentation method of [4] is used. The ground plane is first segmented using the RANSAC algorithm to estimate the 3D plane parameters. From the remaining points, Euclidean clustering is used to form clusters consisting of points that neighbor each other within a margin of 0.15m. As new scan points are acquired, they are matched to the closest existing clusters and each cluster is then updated. New scan points also cause new clusters to be initiated if no neighboring clusters are found. This allows the segmentation process to occur incrementally.

Next, the following relevant entities are automatically labeled in the scene: (i) laser-scanning robot, (ii) construction equipment, and (iii) target assets for manipulation tasks. The position of the laser-scanning robot can be easily determined as the origin of the point cloud scan. On the other hand, the positions of various construction equipment can be determined by using the method in [5], where a feature descriptor is computed for each point cloud cluster and classified with a pre-trained classifier. Finally, the point cloud clusters for objects lying on the ground close to the construction equipment are labelled as potential targets for manipulation.

To perform pose estimation on articulated equipment such as excavators, the strategy used is to approximate each joint as a line segment and use line detection algorithms to extract corresponding segments from the point cloud. Due to the noisy nature of the point cloud, robust parameter estimation methods such as RANSAC and PCA are used. Then the tangent direction of each line segment is used to estimate the joint rotation angles. Empirically, using RANSAC for detecting the boom and PCA for detecting the arm gives the best results. Next, the end effector points are segmented by taking the rightmost points of the excavator after correcting for rotation. The gripper centroid is calculated by taking the centroid of this segment whereas the gripper orientation can be inferred from the internal joint angles. The gripper opening angle is estimated by taking the width of the end effector segment after correcting for rotation. On the other hand, to perform pose estimation of a target asset for manipulation, the position is first estimated by taking the centroid of the corresponding point cloud segment. Then the orientation is estimated by extracting the convex hull and solving for the minimum bounding box. Each step of the pose estimation process also employs error correction using the smooth motion constraint. That is, if a new prediction differs from the previous prediction by more than a predetermined threshold (e.g. due to outlier data), the new prediction is discarded.

The pose estimation results are visualized as shown in Figure 12. The left 4 columns show results for a propane tank as the target object whereas the rightmost column show results for a plastic container as the target object. From the 2D images of the operator view, it is difficult to perceive whether the gripper is in line with the target object and the distance of the gripper to the target object. On the other hand, the 3D visualization provides helpful information in terms of physical distances and poses that can assist the remote operator in decision making.

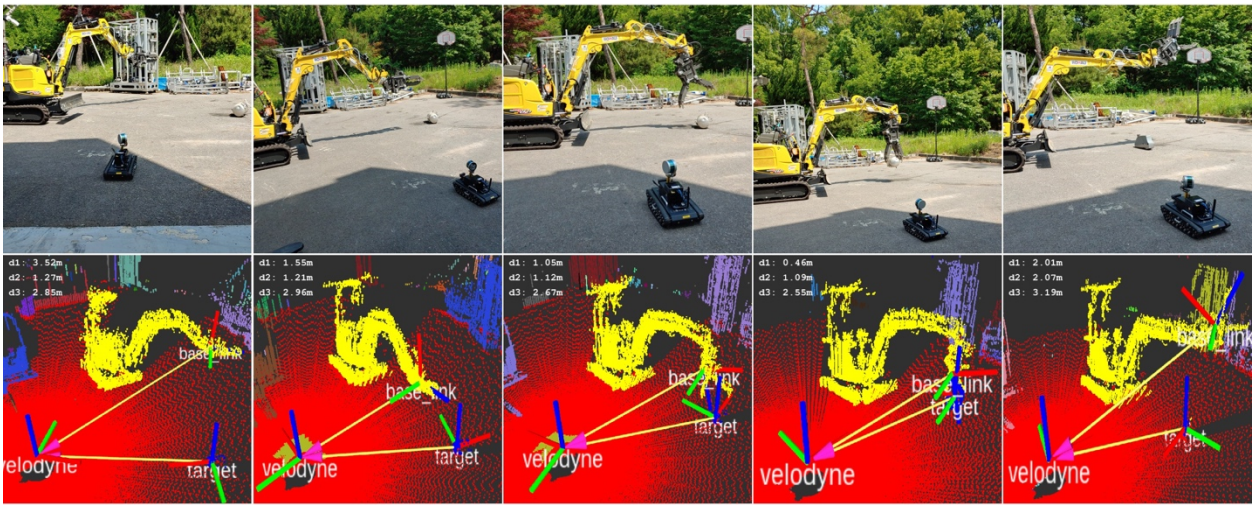


Figure 12: Visualization and pose estimation of excavator workspace. The top row shows 2D images of the scene whereas the bottom row shows the corresponding 3D visualization

Task 3:

Dr. Ueda leads research efforts to accomplish Objective 3. The objective of Dr. Ueda's research group is to establish a shared control framework for a multi-DOF dexterous hand-arm system with adaptive gain scheduling. In the proposal, two research tasks: Task 3-2: Shared Control for Scaled Hand-Arm Tele-operation, and Task 3-1: Adaptive Gain Scheduling for Robust Tele-operation, have been proposed to accomplish Objective 3. Task 3-1 will use outcomes of Objective 1 led by Dr. Cho. Developed control algorithms will be finally implemented onto unmanned excavator robots operated by the GHOST system at the Hanyang University (Objective 3). To interface Objective 3 (Ueda) with Objective 4 (Han), the two research teams discussed (See **Research Collaboration Efforts with the Korean Team** section) and identified a couple of motion control challenges. With these new tasks, subtasks of Object 3 will form a 3-layer structure as shown in Figure 13.

The control system between the human operator using the GHOST system and the remote excavation system is proposed to have three layers:

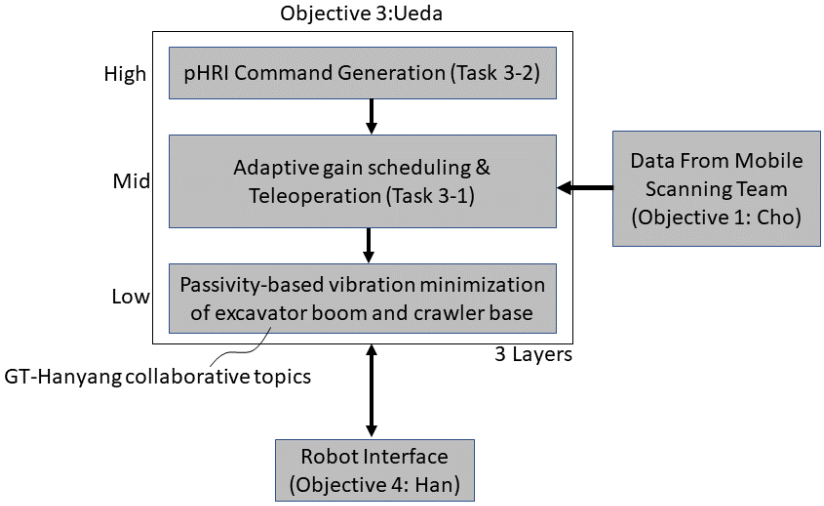


Figure 13: Overview of proposed controller

Each layer corresponds to a control task with descending levels of abstraction and control. The first is a physical human-robot-interaction command generation task. This would include the use of human haptic and bio-feedback information to generate a prediction of the human operator's intention with an associated probability. This information is coupled with data from the other teams on the project along with existing ground information to make a modified control task that compensates for outside risk factors. The final and lowest layer is the motion control of the excavator system, which would include a base-compliance problem for the remote excavator as well as the problem of teleoperation between the GHOST and the remote excavator. The base-compliance problem is approached using posture optimization based on robustness metrics of a nominal excavator model that exploits the end-point positioning redundancy present in an excavation system.

Task 3-1: Adaptive Gain Scheduling for Robust Tele-operation
Low-level (Motion control)

High speed manipulation tasks generate an unavoidable end-effector position error that must be compensated to retain high accuracy. The collocation of sensors to manipulator actuators misrepresents the global position of the end-effector due to an inherent flexibility in the base that is excited by the high-speed motion of the manipulator. Naturally there exist configurations of the manipulator that may be more robust to these disturbances, and the controller for the manipulator may be able to prioritize these configurations over others in some optimal control scheme. However, the compensation provided by the actuators could lead to an instability in the closed-loop control of the end-point position of the end-effector.
Because the GHOST system replaces a human operator and is mechanically coupled to the oscillating system, the effect of feedthrough may overtake the controlled motion of the end-point and induce further end-point instability.
For this, it becomes necessary to be able to model the viscoelastic characteristics of the excavator system used as a remote manipulator for this project. Accelerometers mounted on the cabin, crawler-base, and gripper/end-effector of the excavator fitted with the GHOST telemanipulation system show that for some motion tasks there is base-excitation for the excavator, shown in the figures below. These tasks were performed at high speed and were unloaded, so all excitations were generated using the system's inertia.

Figure 14: IMU Locations on the Remote Excavation System

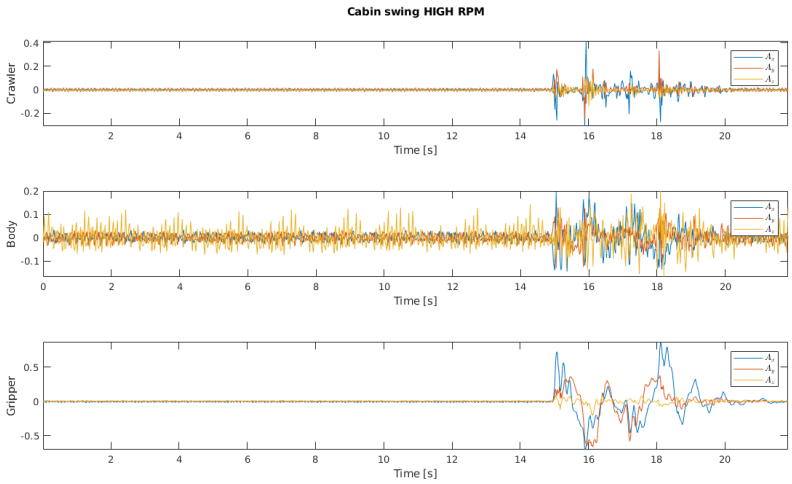


Figure 15: Accelerometer data for high speed spin of the cabin about its center axis

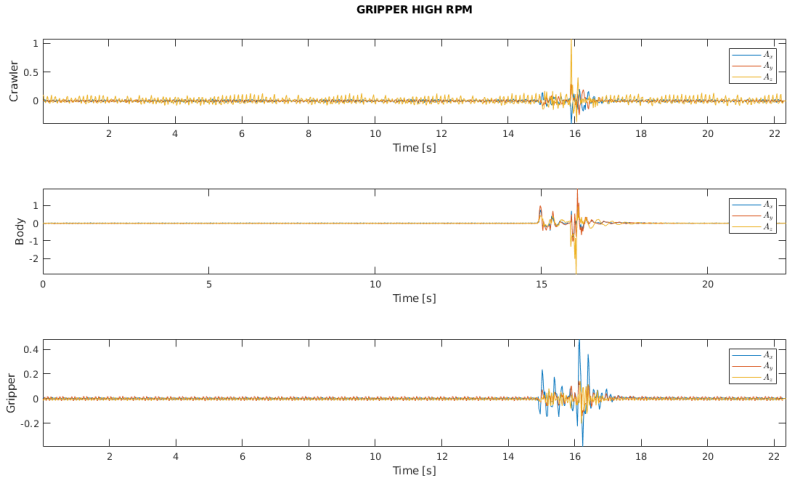


Figure 16: Accelerometer data for high speed lifting of the gripper in the task space

It can be seen from the above data sets that suddenly stopping both the gripper or the cabin from a high-speed motion generates oscillations through the whole excavation system. For the first motion command, shown in Figure 15, the cabin swing, there is significant chatter on the crawler and body axes during the commanded motion (which would be all time before the large peak, about 0 to 15 seconds) which indicates that there is compliance intrinsic to the structure of the excavator. The base-compliance case is supported by the oscillatory behavior shown beginning at about 15 seconds where it can be seen there is significant accelerations along all three axes, as well as an induced oscillation of lower frequency in the gripper. The gripper motion, in Figure 6 above, shows similar behavior where there is a low-amplitude, high frequency oscillation induced in the crawler-base of the excavator and upon stopping the motion command there is significant end-point oscillation in the z-direction (which is some resultant vector normal to the gripper body, shown in Figure 14). This data is significant in highlighting the need for vibration suppression through the entirety of the remote excavation system, and while this issue could be endemic to the small remote unit employed in the test, a case can be made that as the excavator scales in size along with its payload similar effects may be observed.

In making attempts to eliminate the base flexibility problem from the manipulation task, it may become necessary to classify the ground condition at the disaster site. This will necessitate both offline data related to the environment as well as the performance of the excavation system on that ground condition.

This would imply that the compensation methods used must have a strong initial data set that may be informed by the numerical model and is adapted using information from the sensor systems on-board the excavator. This data can be supplemented by information from the mobile scanning team about the ground content for the area within a region achievable by the motion of the excavator.

Applying robust control methods based on a linearized excavator plant about a single arm posture, $P(s, \boldsymbol{q})$, end-point positioning performance can be improved for the mobile robotic system where $\boldsymbol{q}$ is the joint angle vector of the manipulator. Robustness of the end-point control can be characterized by the passivity of the excavator with oscillatory base seen from an end-point position control scheme. Two robustness metrics are known to perform almost equivalently and are introduced.

The first method applies coprime factorization of a linearized plant and then provide a metric based on controller robust loop shaping to describe the robustness of a posture of linearization to modeling error. This metric, ε_{max}, characterizes the robustness of the posture for all stabilizing controllers applicable to the system and hence is not dependent on the kind of controller used except for the choice of task-space controller, either Jacobian inverse or Jacobian transpose, that are popular control methods. Mathematically, the robustness metric is given by:

$$\widetilde{\boldsymbol{P}}(s) \triangleq k_r \boldsymbol{P}(s) = \widetilde{\boldsymbol{M}}^{-1}\widetilde{\boldsymbol{N}}$$
$$\varepsilon_{max} \triangleq \sqrt{1 - \left\| [\widetilde{\boldsymbol{N}} | \widetilde{\boldsymbol{M}}] \right\|_H^2}$$

Where, $k_r > 0$, is a servo gain used to perform the analysis with a higher gain. $\|G\|_H$ is the Hankel norm of G. For this metric, robustness is large when ε_{max} is large. Examples of these surfaces are shown in Figure 17.

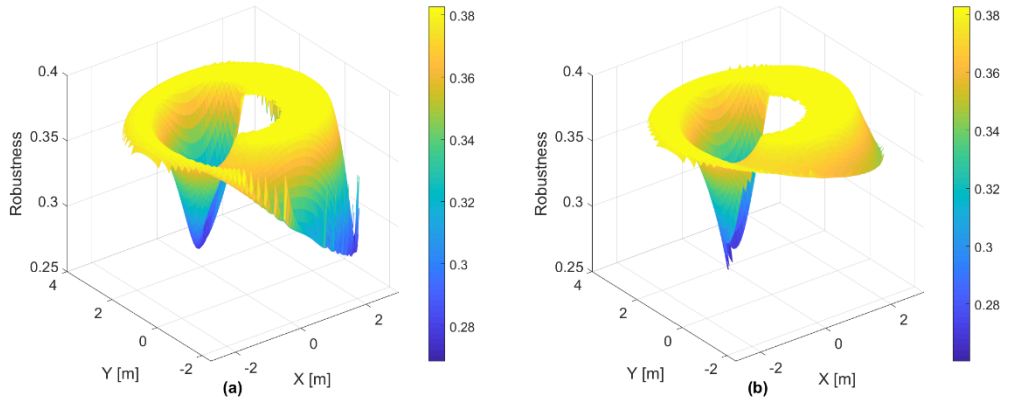


Figure 17: ε_{max} robustness metrics for (a) Jacobian transpose and (b) Jacobian inverse control

The second metric uses modal decomposition of the linearized plant ($\boldsymbol{P}(s, \boldsymbol{q}) = \boldsymbol{C}(sI - \boldsymbol{A})^{-1}\boldsymbol{B}$) to identify vibrational modes of the system. Using Jordan decomposition of the plant state matrix, $\boldsymbol{A}(\boldsymbol{q})$, a matrix containing generalized eigenvectors, $\boldsymbol{U}$ (and its inverse $\boldsymbol{V}^*$), can be found. The residue matrix of the i^{th} vibrational mode is

$$R_i = -2 * Re(\bar{\lambda}_i \boldsymbol{C}\boldsymbol{u}_{2(n+i)-1}\boldsymbol{v}^*_{2(n+i)-1}\boldsymbol{B})$$

Where $\bar{\lambda}_i$ is the conjugate eigenvector of the i^{th} vibrational mode, $\boldsymbol{u}_x$ is the is the x^{th} row of $\boldsymbol{U}$, $\boldsymbol{v}_x^*$ is the is the x^{th} column of $\boldsymbol{V}^*$, and n is the number of rigid body modes in the system. This residue matrix can be used to make a robustness metric w_i associated with the i^{th} vibrational mode of the system.

$$w_i = \lambda_{\min}\left(R_i + R_i^T\right) \leq 0$$

This metric indicates the effective distance of a posture from a *robust arm configuration*, a posture wherein the vibrational mode is in-phase and passive. The resulting w_i metric is shown in Figure 18. Values closer to zero are considered to achieve greater robustness.

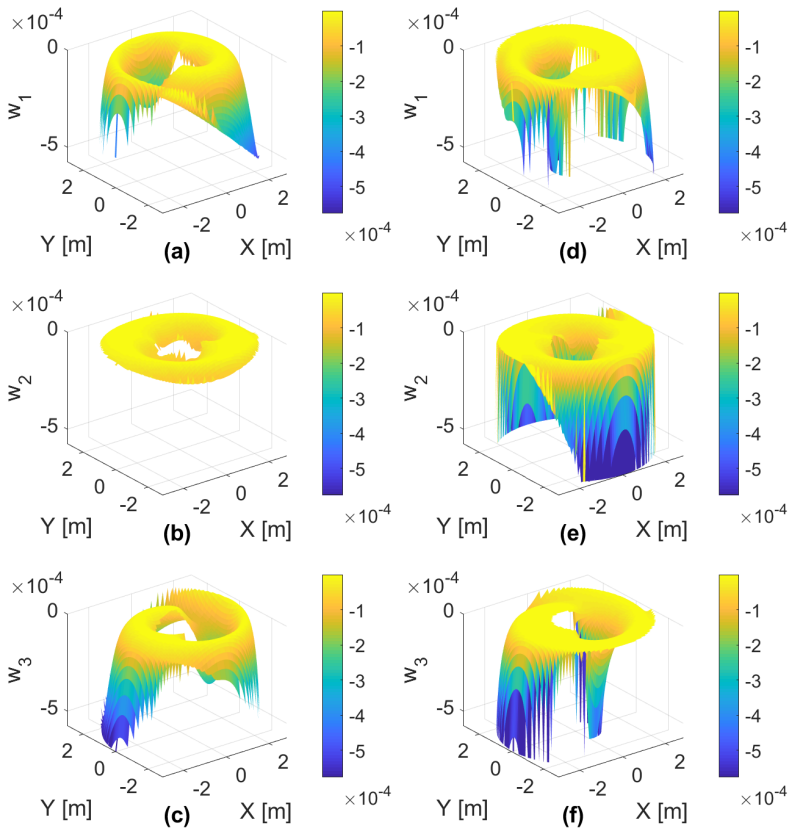


Figure 18: w_i, modal decomposition-based robustness for (a)-(c) Jacobian transpose control and (d)-(f) Jacobian inverse control

Using these two metrics, one can apply nullspace augmentation of the nominal posture of the manipulator according to the gradient of one of these metrics.

$$\Delta q_d = (I - \boldsymbol{J}_{\boldsymbol{B}}^{+}\boldsymbol{J}_{\boldsymbol{B}})k_\xi \xi$$

Here, Δq_d is the displacement of the nominal joint states for the manipulator posture, I is an identity matrix, $\boldsymbol{J}_{\boldsymbol{B}}$ is the Jacobian relating the end-point velocity to the joint states (considering the redundancy of the system), k_ξ is a scalar gain used to damp the result, and ξ is the gradient of the objective function that is to be maximized.

This method projects the objective function across the nullspace of the redundancy of the manipulator to augment the nominal states of the manipulator. In the case of a remote planar excavator, the redundancy

of the system is in the positioning of the manipulator's mobile base. Using this it becomes possible to change the posture of the manipulator without changing the position of the manipulator end-point in the inertial, world frame.

Base Oscillation Minimization Theory

Defining the remote, planar excavator as a 2DOF planar manipulator with a 3DOF planar flexible base, shown in Figure 19, the system can be analyzed using the methods described above. Here, the system has two rigid modes (one for each link in the manipulator) and three vibrational modes ($n = 2, m = 3$).

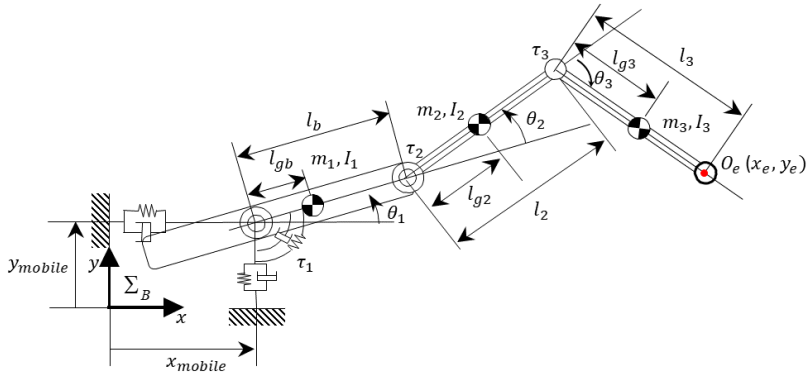


Figure 19: Planar manipulator with flexible base, end-point shown in red.

In many practical applications, manipulation systems are assumed to be rigid with resonant frequencies significantly higher than the control bandwidth. For linearization and control system generation, the flexible system is assumed to be rigid to examine the stability and complexity of the feedback control. The resulting system behaves like the manipulator shown in Figure 20.

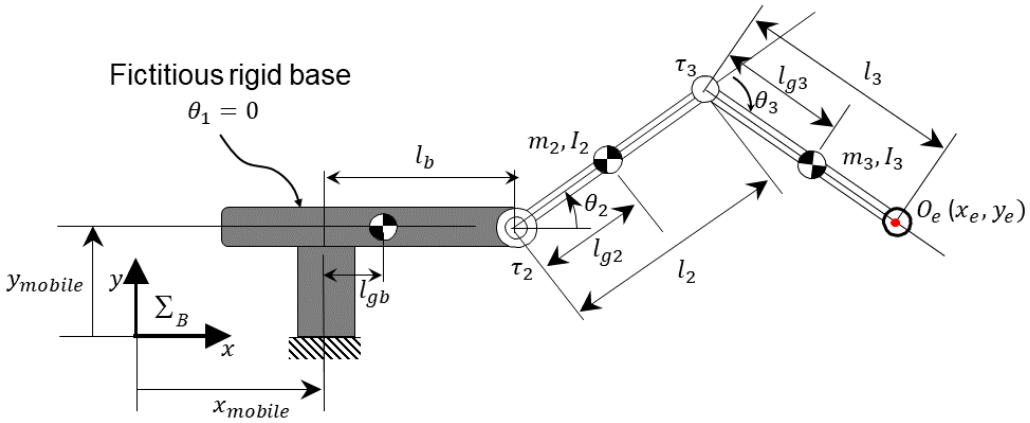


Figure 20: Manipulator model used in feedback control design

The robust control methods described above are applied a nonlinear simulation of the planar remote excavation system using parameters given in Table 1.

Name	Value	Unit	Name	Value	Unit
l_b	0.541	m	l_{gb}	0.563	m
l_1	1.918	m	l_{g1}	1.067	m
l_2	1.097	m	l_{g2}	0.613	m
m_b	5291.5	kg	I_b	1020.7	kgm^2
m_1	766.6	kg	I_1	214.9	kgm^2
m_2	458.2	kg	I_2	93.6	kgm^2
k_x	$4x10^7$	Nm^{-1}	b_x	50	Nsm^{-1}
k_y	$4x10^7$	Nm^{-1}	b_y	50	Nsm^{-1}
k_θ	$1.112x10^6$	Nm^{-1}	b_θ	25	Nsm^{-1}
θ_{1d}	1.2451 rad				

Table 1: Dynamic system parameters for the excavator

The ε_{max} and w_i metrics are calculated across the workspace of the manipulator, shown above in Figures 17 and 18. It can be seen between the postures listed in Table 2 that there is an increased performance at the posture of better robustness for the same controllers from the simulation data given in Figures 22 and 23.

Posture	A	B
$x_{desired}$	0.00m (inertial) −0.1733m (workspace)*	0.00m (inertial) 1.32687m (workspace)*
$y_{desired}$	1.60m (inertial) 1.0870m (workspace)*	1.60m (inertial) 1.0870m (workspace)*
x_{mobile}	0.00m	−1.50m
ε_{max}	0.2869	0.3826
w_1	-0.2969×10^{-3}	-0.0209×10^{-4}
w_2	-0.0308×10^{-3}	-0.0683×10^{-4}
w_3	-0.001×10^{-3}	-0.3105×10^{-4}

Table 2: Chosen postures for stability study

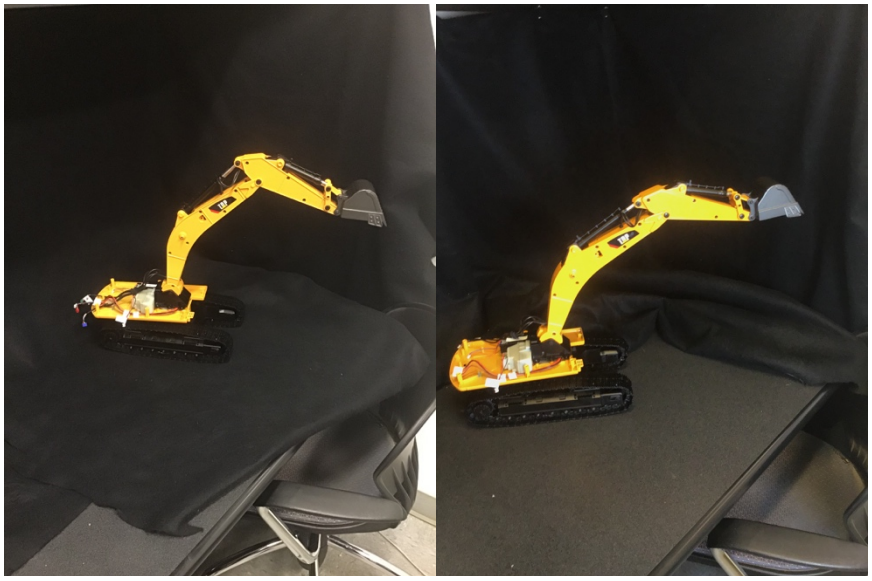


Figure 21: Scaled excavator in postures analogous to those in Table 2.

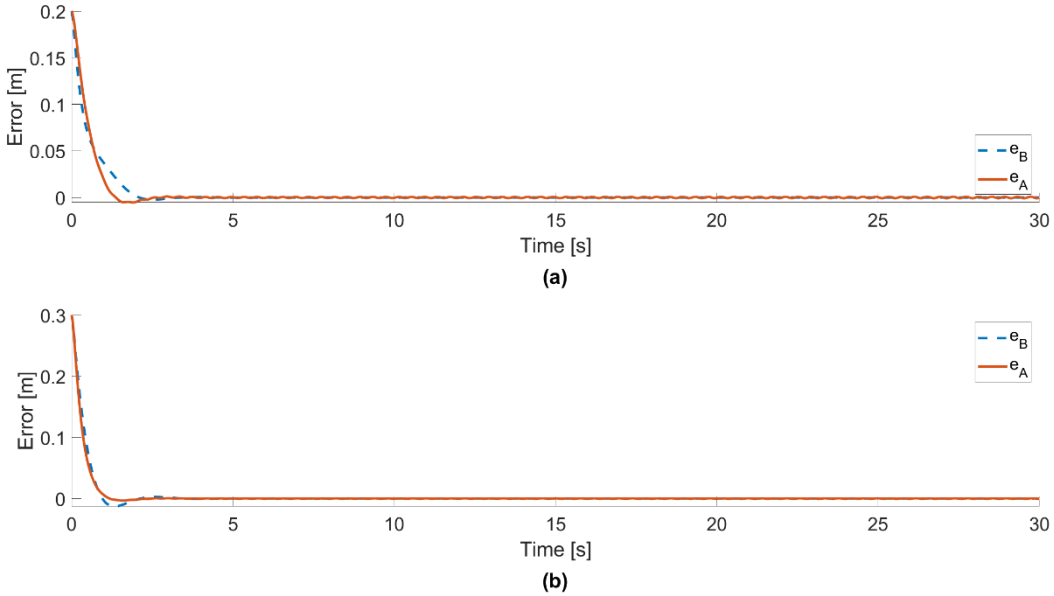


Figure 22: End-point position error for the chosen optimization positions using Jacobian inverse controller. (a) is x axis error and (b) is y-axis error.

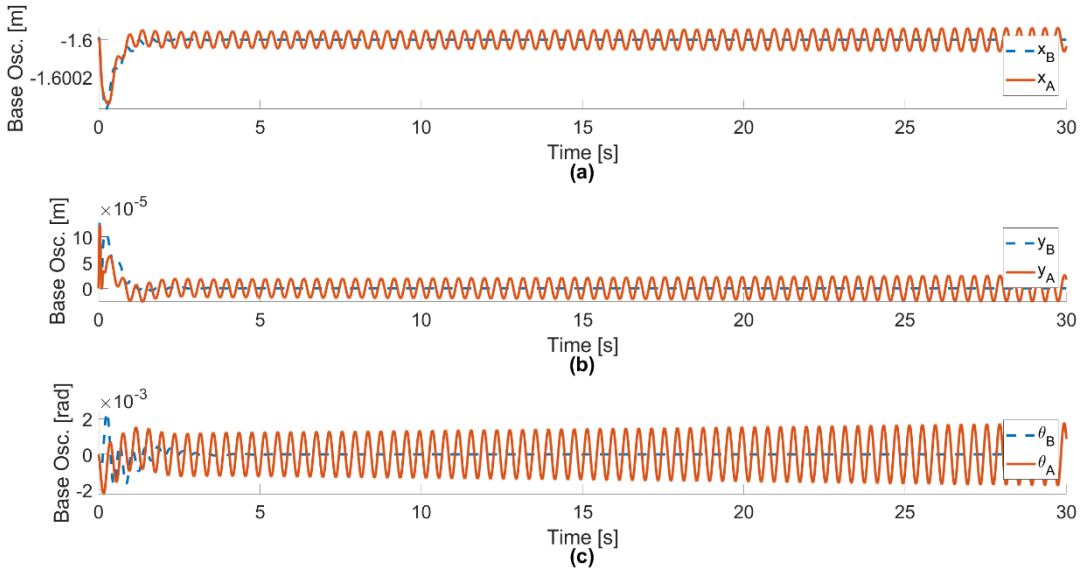


Figure 23: Base oscillations for the chosen optimization positions using Jacobian inverse controller. (a) is x axis oscillation, (b) is y axis oscillation, and (c) is the angular oscillation of the manipulator base.

Taking slices of each robustness metric surface along the chosen $y_{desired}$ it becomes possible to achieve an optimization curve to best position the manipulator to have an optimal posture for the chosen end-point. A curve taken along the $y_{desired}$ given in Table 2 is shown in Figure 24.

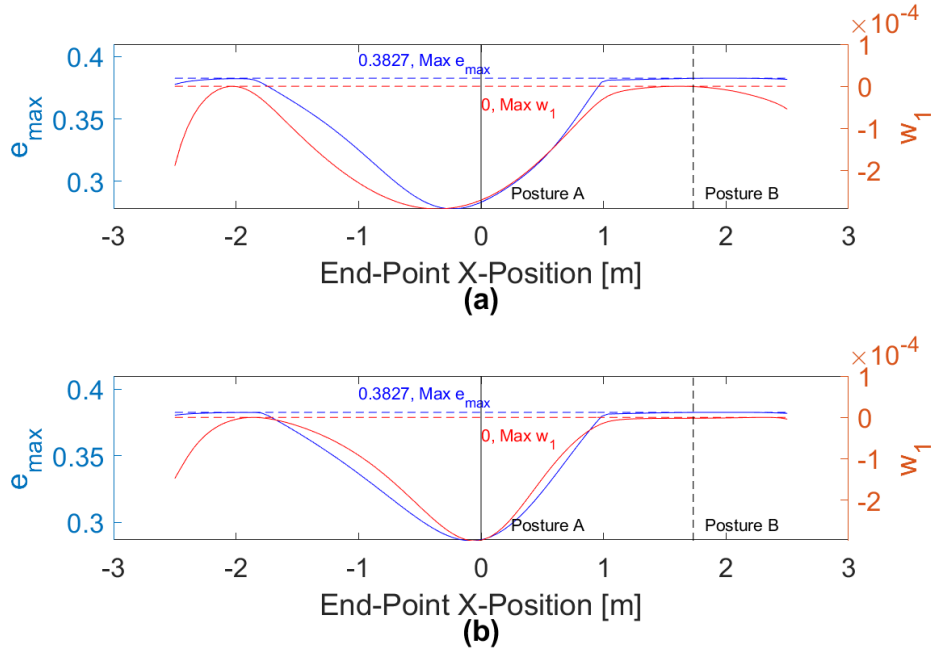


Figure 24: Optimization curves for (a) Jacobian transpose control and (b) Jacobian inverse control along $y_{desired}$. Due to the presence of the second and third vibrational modes, ε_{max} is always less than its maximum of 0.3827.

An objective function for maximizing modal robustness in the task-space of the manipulator is generated using a linear combination of the w_i values for the system.

$$V(\boldsymbol{q_d}) = \sum_{i=1}^{m} \alpha_i w_i(\boldsymbol{q_d}), \alpha_i = \{0,1\}$$

The use of the α_i coefficient is used to turn each robustness "on" or "off" for consideration in the objective function. To use this objective function with the presented nullspace augmentation method, its gradient with respect to the redundant joint states ($\boldsymbol{q_B} = [x_{mobile}, \theta_2, \theta_3]^T$), ξ, is required.

$$\xi = \frac{\partial V(\boldsymbol{q_d})}{\partial \boldsymbol{q_B^T}} = \sum_{i=1}^{m} \alpha_i \frac{\partial w_i(\boldsymbol{q_d})}{\partial \boldsymbol{q_B^T}}$$

The gradient along a particular parameter, q_l, of a modal robustness w_i is given as:

$$\frac{\partial w_i(\boldsymbol{q})}{\partial q_l} = \sum_{a=1}^{n} \left(\sum_{b=1}^{n} \left(\frac{\partial \left[R_i + R_i^T\right]_{ab}}{\partial q_l} \nu_b \right) \nu_a \right)$$

Where $\left[R_i + R_i^T\right]_{ab}$ is the element in the a^{th} row and b^{th} column of the matrix $\left[R_i + R_i^T\right]$; ν is the eigenvector associated with $\lambda_{min}\left(R_i + R_i^T\right)$, and ν_x is the x^{th} element in ν.

Using the nullspace augmentation of the nominal joint states to shift the posture of the manipulator from Posture A to Posture B, one can use the gradient $\frac{\partial V(q_d)}{\partial x_{mobile}}$ as a velocity to exploit the redundancy of the manipulator's mobile base. The resulting trajectory is shown below in Figure 25.

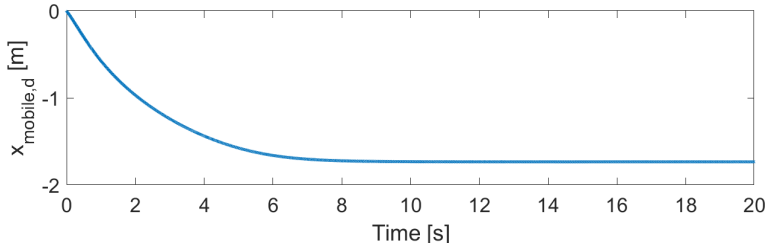


Figure 25: Base position trajectory for manipulator using nullspace augmentation.

The simulation results comparing end-point positioning tasks with and without base motion are shown in Figures 26 and 27 below.

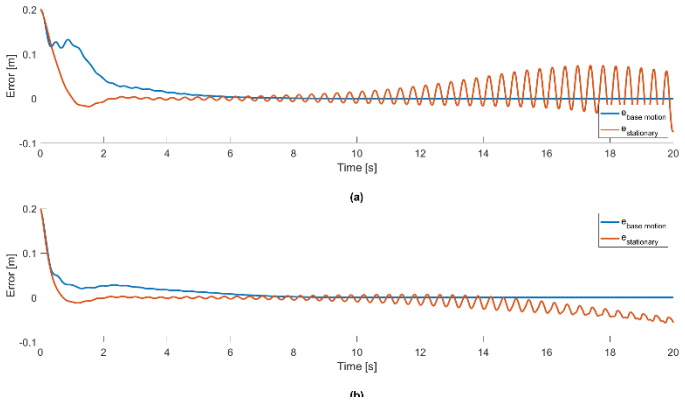


Figure 26: End-point position error with and without base position adjustment using Jacobian inverse controller. (a) x-axis error, (b) y-axis error. With base position adjustment changing from posture A to posture B using the base trajectory shown in Figure 14 during end-point positioning (shown in blue). Without base position adjustment (shown in red) endpoint positioning around Posture A.

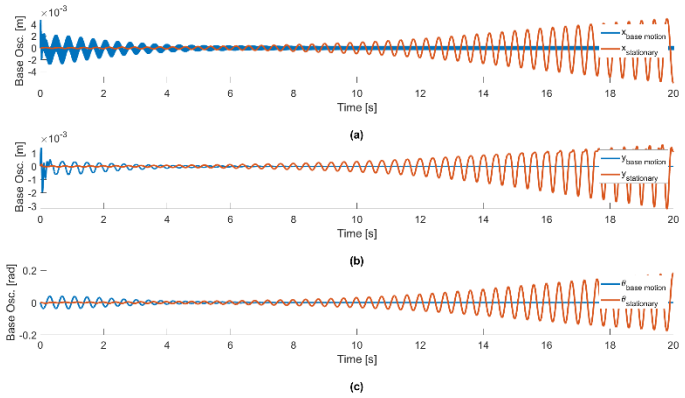


Figure 27: Base oscillation with and without base position adjustment using Jacobian inverse controller. (a) x_{mobile} oscillation, (b) y_{mobile} oscillation, and (c) θ_1 oscillation. With base position adjustment changing from posture A to posture B using the base trajectory shown in Figure 19 during end-point positioning (shown in blue). Without base position adjustment (shown in red) end-point positioning around Posture A.

Proof of Stabilizability

The work conducted this reporting period focused on an investigative study into the properties of the stability margins used to optimize the posture of the manipulator against the unknown base disturbances. The theory leveraged in the base optimization of the previous section can be summarized as:

> If a rigid manipulator on a flexible base is in a posture where its flexibilities are completely passive (this posture is known as an RAC, robust arm configuration; all modal-space stability indices are zero, $w_i = 0 \; \forall i = 1, \dots, m$), then the value of ε_{max} is maximized and is equivalent to that of a rigid manipulator on a rigid base.

The above can be broken into:

1. The posture must be an RAC ($w_i = 0 \; \forall i = 1, \dots, \#FlexibleModes$)
2. At this posture the value of ε_{max} is maximized
3. This maximum value is the same as that of the rigid system

Satisfaction of the first item in the above list requires the localization of RAC's within the manipulator workspace. Defining the modal-space stability index as

$$w_i = \lambda_{min}\left(R_i + R_i^T\right) \leq 0$$
$$R_i = -2Re[\bar{\mu}_i C O_i B] = R_i^T$$

Where R_i is the residual matrix associated with the i^{th} flexible mode, $\bar{\mu}_i$ is the conjugate of the eigenvalue of the i^{th} flexible mode, C and B are the state-to-output and input-to-state matrices from a joint-space state-space model of the manipulator system, and O_i is the spectral operator associated with the i^{th} flexible mode.

While the second point of the theory above is complex and still being analyzed, the third point is complete. In MacFarlane's lectures, the definition of the ε_{max} stability margin is defined as

$$\varepsilon_{max} = \frac{1}{\sqrt{1 + \lambda_{max}(ZX)}}$$

Where Z and X are the unique, Hermitian, positive definite solutions to the Generalized Filtering Algebraic Riccati Equation (GFARE) and the Generalized Controller Algebraic Riccati Equation (GCARE) [15]. This provides a means to generate analytical solutions for the ε_{max} stability margin.

Robust Arm Configuration Conditions

Evaluating $w_i = 0$ for the general manipulator joint-space state-space equations, it was found that this occurs when either the C or B matrix is rank-deficient. Practically this would mean that the flexible modes of the flexible-base manipulator system are either unobservable or uncontrollable. This is ideal given the nature of the system because it implies passivity of the flexible modes.

Maximum $\boldsymbol{\varepsilon_{max}}$ Stability Margin

Given a general rigid-base manipulator model:

$$A = \begin{bmatrix} 0 & I \\ 0 & 0 \end{bmatrix}, B = \begin{bmatrix} 0 \\ PR_0 \end{bmatrix}, C = \begin{bmatrix} P^{-1} & 0 \end{bmatrix}, D = 0$$

Where R_0 is the rigid mode of the manipulator system and the invertible matrix P is used to decompose the rigid mode between the input and output sides of the system.

Solving both the GCARE and GFARE for a rigid-base manipulator state-space model yields a constant matrix-product for (ZX) in the ε_{max} definition. This product is

$$Z = \begin{bmatrix} \sqrt{2}P(R_0R_0^T)^{\frac{1}{4}}P^T & P\sqrt{R_0R_0^T}P^T \\ P\sqrt{R_0R_0^T}P^T & \sqrt{2}P(R_0R_0^T)^{\frac{3}{4}}P^T \end{bmatrix}, X = \begin{bmatrix} \sqrt{2}P^{-T}(R_0R_0^T)^{-\frac{1}{4}}P^{-1} & P^{-T}\left(\sqrt{R_0R_0^T}\right)^{-1}P^{-1} \\ P^{-T}\left(\sqrt{R_0R_0^T}\right)^{-1}P^{-1} & P^{-T}\sqrt{2}(R_0R_0^T)^{-\frac{3}{4}}P^{-1} \end{bmatrix}$$

$$(ZX) = \begin{bmatrix} 3 & 2\sqrt{2}P(R_0R_0^T)^{-\frac{1}{4}}P^{-1} \\ 2\sqrt{2}P(R_0R_0^T)^{\frac{1}{4}}P^{-1} & 3 \end{bmatrix}$$

$$p_{ZX}(\lambda) = \lambda^2 + 6\lambda + 1 = 0$$

$$\lambda = 3 \pm 2\sqrt{2}$$

$$\varepsilon_{max} = \frac{1}{\sqrt{4 + 2\sqrt{2}}}$$

The maximum eigenvalue of the matrix product is $3 + 2\sqrt{2}$, which leads to a maximal value of ε_{max} of ~ 0.38268. This is the point of comparison for the flexible-base manipulator as it implies that the flexibilities present in the system are dissipative in nature. It should be noted that the characteristic

equation, $p_{ZX}(\lambda)$ is not dependent on the rigid mode of the system. This implies that this maximum value should hold for all MIMO inertial systems.

Once the manipulator is at a robust posture, conditions to remain in that posture have been deduced. Considering the hermite matrix $H_i = R_i + R_i^T$, the matrix used to calculate the w_i stability index, one can define a perturbation to that matrix, ΔH_i. At a robust posture, R_i is symmetric, making $\Delta H_i = 2\Delta R_i$. When the system is perturbed away from its robust condition in the workspace, but remains robust the minimum eigenvalue of the hermite matrix should remain at 0. Looking at perturbation of the eigenvalue problem for the hermite matrix:

$$H_i v_j = \lambda_j v_j$$
$$(H_i + \Delta H_i)(v_j + \delta v_j) = (\lambda_j + \delta\lambda_j)(v_j + \delta v_j)$$

Where λ_j and v_j are the j^{th}eigenvalue and its associated eigenvector for the hermite matrix H_i. These are perturbed by some small δ each. The problem can be expanded with higher-order terms removed to achieve:

$$\Delta H_i v_j + H_i \delta v_j = \delta\lambda_j v_j + \lambda_j \delta v_j$$

The eigenvector perturbation can be represented by a linear combination of existing orthogonal eigenvectors. Making this substitution, the equation can be represented as

$$(\lambda_j + \delta\lambda_j) = \lambda_j + \frac{v_j^T \Delta H_i v_j}{v_j^T v_j}$$

The above holds for systems of any dimension but looking at the conditions for a planar manipulator system. The zero eigenvalue of the hermite matrix must remain at zero, substituting in $\lambda_1 = 0$ with its associated eigenvector v_1

$$(0 + \delta\lambda_1) = \frac{v_1^T \Delta H_i v_1}{v_1^T v_1} = 0$$
$$\delta\lambda_1 = \frac{v_1^T \Delta H_i v_1}{v_1^T v_1} = 0$$

This yields the condition that $v_1^T \Delta H_i v_1 = 0$. Analyzing the second, positive eigenvalue for the hermite matrix of the planar system

$$(\lambda_2 + \delta\lambda_2) = \lambda_2 + \frac{v_2^T \Delta H_i v_2}{v_2^T v_2} \geq 0$$
$$-\lambda_2 \leq \frac{v_2^T \Delta H_i v_2}{v_2^T v_2}$$
$$-\lambda_2 \leq \frac{\lambda_{min}(\Delta H_i)\|v_2\|^2}{\|v_2\|^2} \leq \frac{v_2^T \Delta H_i v_2}{v_2^T v_2}$$
$$-\lambda_2 \leq \lambda_{min}(\Delta H_i)$$

This condition is always valid because the ΔH_i must be at least positive semi-definite so that the new hermite matrix after its perturbation remains positive semi-definite. For this, the only condition that contributes is the first, requiring $v_1^T \Delta H_i v_1 = 0$.

Experimental Setup

In order to test the theory developed in the past year of the project, a scale-model of an excavator is being outfit to be a robotic device. The system will utilize modular design, so the sensor systems and

control theory-programming can be swapped between different mechanical and computer systems with minimal necessary changes. Human-interface will be done with a 3D mouse (3D Connexion, SpaceNavigator) to provide motion commands to the mobile excavation system. The main computer is a single chip computer running a Linux distribution and interfaces the sensor and motor drive systems using USB.

Mid-level (Adaptive gain scheduling and teleoperation control

Teleoperation control has a goal in transmitting force and position between the master and slave robots without any intervening impedance originating within the robotic system itself. This is to say that the operator should be able to affect his environment using the robotic system as if he was directly manipulating it. The system should be transparent.

Figure 28: Ideal case of teleoperation, a massless magic hand

This requirement imparts the following conditions on the system:

$$f_m = -f_s$$
$$x_m = x_s$$

The proposed controller will employ a form of oblique coordinate control. Oblique coordinate control is a coordinate transformation used to orthogonalize tasks that do not behave on the same axes. This transformation is conducted by means of a task-Jacobian which maps the natural-coordinate position and force vectors for the master and slave robots into position and force commands (specifically the position command for the force task and the force command for the position task) for use in generating the actuator commands that will eventually be sent to the master and slave robots respectively.
To compensate for the size difference between the master and slave robot the task-Jacobian can be modified to have the form:

$$J_t = \begin{bmatrix} 1 & -\alpha \\ 1 & \beta \end{bmatrix}$$

Which translates to a scaling factor between the position command

$$x_m - \alpha x_s = 0$$

and the force command

$$f_m + \beta f_s = 0$$

where the generic, matched teleoperation task has α and β both equal to 1. So that the given actuator command has the necessary proportionality for proper force and position feedback to the operator.

While the goal of oblique coordinate control is to orthogonalize control tasks so they act on the same axes it is important to note that energy methods will no longer apply to the dynamics of the system once they are brought into the oblique-modal space. For this it becomes necessary to validate the passivity of the bilateral controller as it returns to the task space from the modal space to verify its stability long term stability. This is done by a singular value analysis of the scattering matrix for the controller, assuming the environment and operator dynamics are strictly passive and non-interactive. The structured singular value for reciprocal scattering matrices lies between the maximum and minimum singular values and is a metric of how perturbations of unknown structure and magnitude affect the dissipativeness of a network.

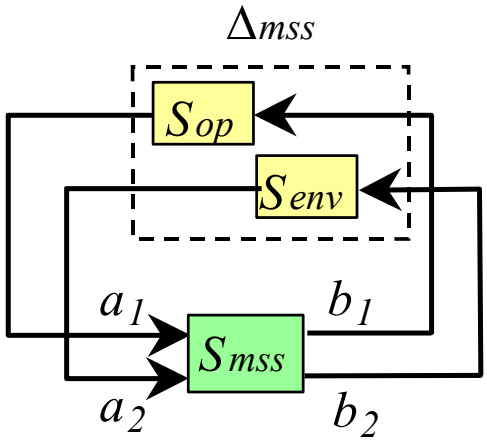


Figure 29: Scattering theory block representation for a generic two port network. Δ_{mss} represents the perturbation of unknown structure to the system generated by the operator and environment dynamics

Applying the singular value analysis to a previously designed modal-space controller:

$$\tau = \frac{1}{2}\begin{bmatrix} -K_p(x_m - x_s) - K_v(\dot{x}_m - \dot{x}_s) - K_f(f_m - f_s) \\ K_p(x_m - x_s) + K_v(\dot{x}_m - \dot{x}_s) - K_f(f_m - f_s) \end{bmatrix}$$

We found that the controller is not guaranteed to be passive in the task space by the decibel magnitude of the maximum singular value, shown in Figure 30, of the scattering matrix being positive for some bandwidth of frequencies. This implies, conservatively, that stability of the control system is not ensured. The ideal behavior of the controller should follow more closely to the passivity-based position and force reflection servoing controller shown in Figure 31.

$$\tau = \begin{bmatrix} -K_p(x_m - x_s) - K_v(\dot{x}_m - \dot{x}_s) - K_f(f_s - f_m) - f_s \\ K_p(x_m - x_s) + K_v(\dot{x}_m - \dot{x}_s) + K_f(f_m - f_s) + f_m \end{bmatrix}$$

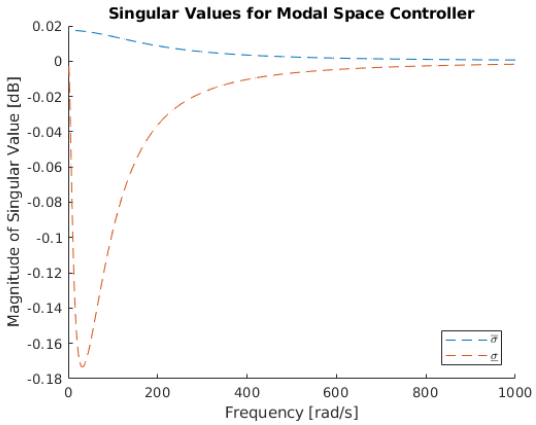


Figure 30: Singular Value Analysis for Modal Space-Derived Controller

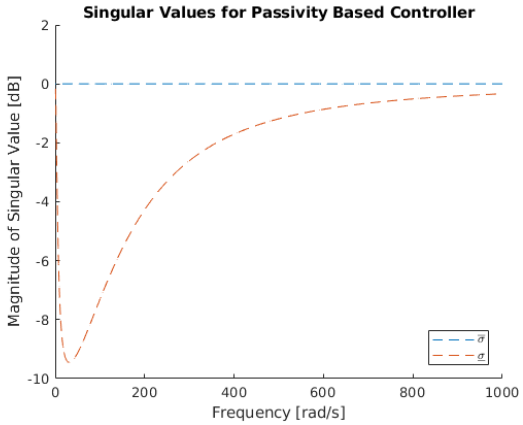


Figure 31: Singular Value Analysis for Passivity-Based Bilateral Controller

While the overall structure of the modal-space controller is similar to the passivity-based bilateral controller, the two largely differ in the force-reflecting servo terms wherein the modal-space controller was designed for proportional control on the common-mode of the forces whereas the passivity based controller was designed with the intention of guaranteeing stability for a remote teleoperation task. This opens the avenue for research into the application of an oblique coordinate transformation in remote bilateral controllers. The teleoperative control scheme is shown in Figure 32.

This analysis was also extended to scaled task jacobians and scaled mass systems.

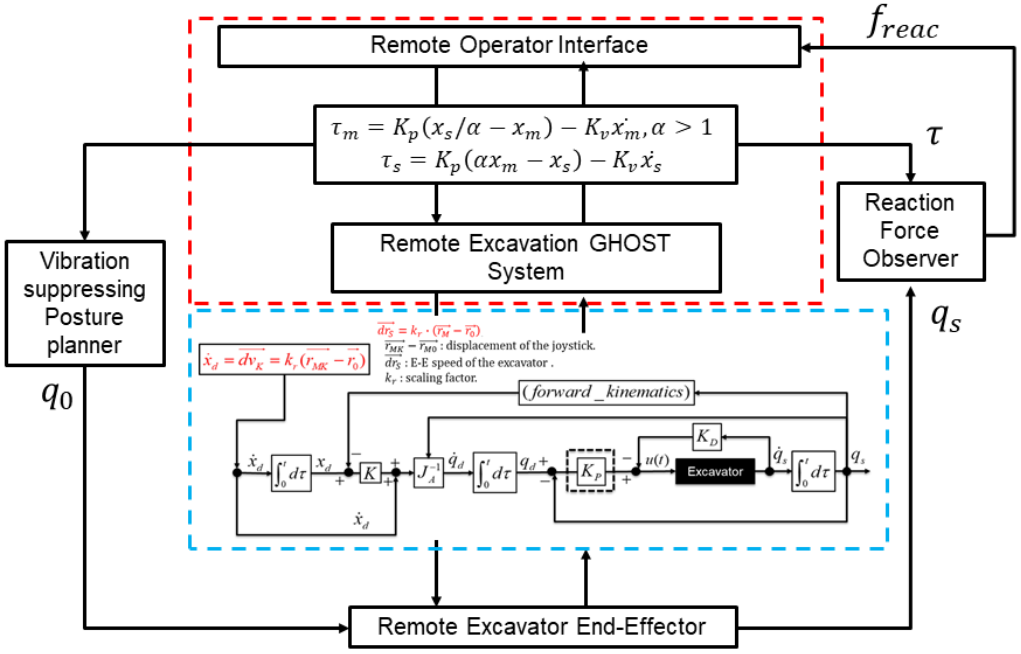


Figure 32: Teleoperative control scheme showing included low-level control in the loop.

The material listed in the above section can be used for the gain-scheduling adaptive control. It has been shown that postures of low stability margin do not reject induced vibrations at high controller gain. It is possible to vary the overall controller gain as the posture of the manipulator changes. This allows for higher gain in postures that can tolerate it without instability.

In addition to the development of stable teleoperation control schemes, the project will also establish control theoretic methods to enhance cyber security of motion control systems. The proposed approach is to encrypt motion control algorithms, feedback and feedforward gains, and sensor data, and perform necessary computation to generate motion commands to servo systems in the ciphertext space by taking the advantage of homomorphic encryption. The proposed encrypted motion control configures motion control systems to improve robustness against various types of cyberattacks and prevent sensor data breach and system instability due to controller falsification. The remote robotic manipulator system benefits from this improved cyber security as shown in Figure 33. Research tasks include (1) development of control architecture better suited for homomorphic computation, (2) characterization of quantization errors, time delays and necessary processing resources, (3) implementation of real-time control software and hardware with controller encryption.

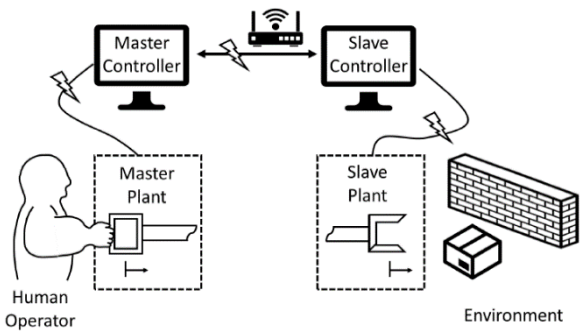


Figure 33: Master-slave teleoperation system.

Lightning symbols indicate potential cyber-physical attacks

This aim proposed a realization of an encrypted masterslave teleoperation system applying ElGamal encryption as shown in Figure 34. A data sharing scheme was provided for a two-key two-end-device (master and slave) system, and the computation of system inputs in ciphertext has been formulated. A security-enhanced control framework is to change the order of the control scheme by introducing stable pole-zero cancellation technique. Simulation results demonstrated that encrypted controller and increased-order encrypted controller could be implemented to enhance system security without compromising system performance.

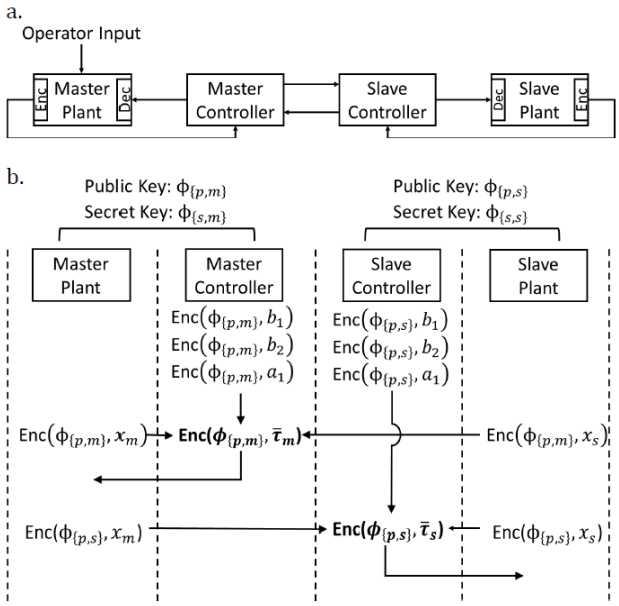


Figure 34: Encrypted teleoperation system with symmetric position control. (a) System components: a master plant, a master controller, a slave controller, and a slave plant. The arrows indicate the routing of encrypted information within the system. (b) System parameters and data flow within the system.

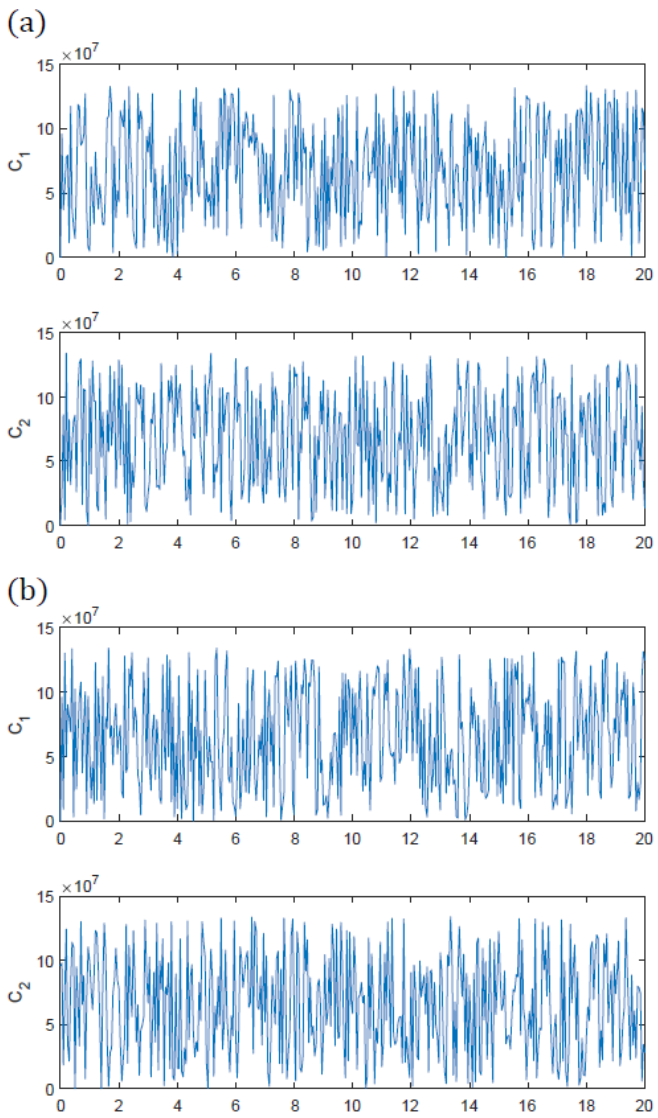


Figure 35: Encrypted teleoperation commands during positioning tasks. (a) encrypted master commands, (b) encrypted slave commands

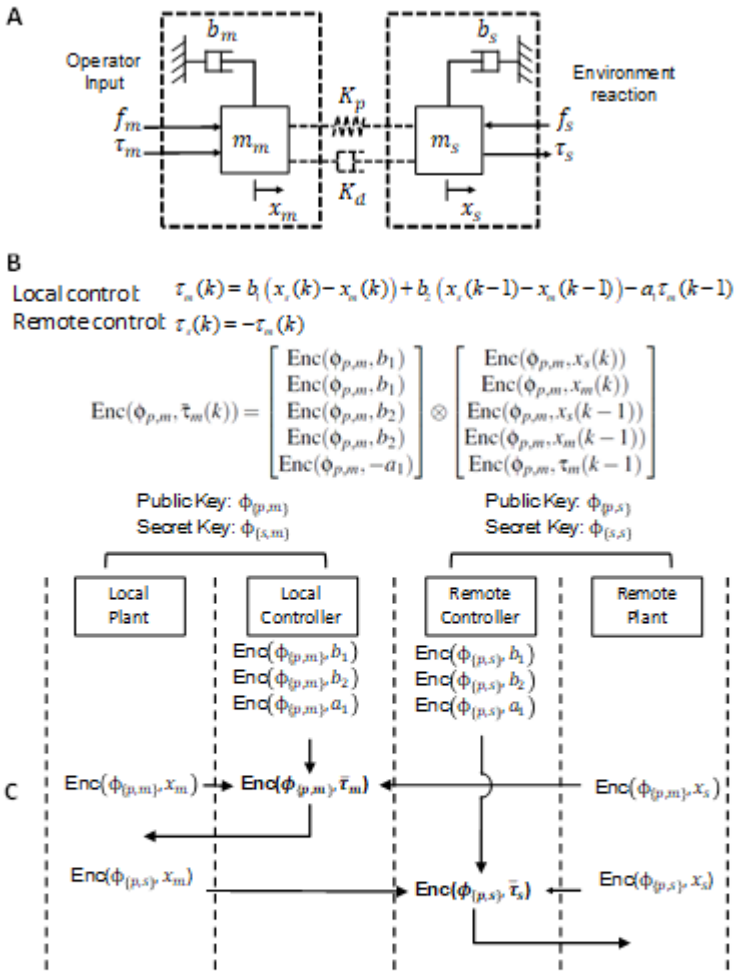


Figure 36: Proposed Implementation of encrypted teleoperation. A. Target symmetric position servo control, B. Discrete-time control schemes for local and remote controllers Elements-wise multiplications in ciphertext-space. C. Elgamal encryption with two sets of keys.

In order to share information and perform computation in the teleoperated motion control system, shared data is encrypted for each end-device, i.e., local and remote sensor data, and controller gains are encrypted into two copies: one encrypted by the local plant's key and the other one by the remote plant's key. The research of this project is to develop a theoretical framework for the implementation of secure encrypted control onto a networked teleoperation system where local and remote robot systems are located remotely and connected to different intra-networks.

Figure 36 shows the concept of encrypted teleoperation system control [6]. Widely-used symmetric position servo control is adopted for initial design that applies proportional-derivative control to both plants to introduce intervening impedance (i.e., virtual spring and damper) between two linear motion systems (Fig. 36A). In the ciphertext space, multiplications compute discretized control schemes and produce intermediate commands (Fig. 36B). Additions are then performed for decoded intermediate commands. Fig. 36C shows system parameters and data flow of each component with a specified encryption key. The main concept here is to encrypt shared information with multiple public keys assigned to the devices that need to perform decryption. For a teleoperation system, it is reasonable to assume that at least two different sets of keys are required: local keys (a public key and a secret key) shared by the local plant and local controller, and remote keys (with a public key and a secret

key) shared by the remote controller and remote plant. Both local and remote plants are responsible for system output measurement and encryption by using both local and remote public keys. The local controller stores system parameters a_1, b_1, and b_2 encrypted with a local public key. It also receives output measurements encrypted with a local public key from both plants. Upon receiving plant outputs, intermediate commands of the local plant input is computed. Since ElGamal encryption does not support ciphertext addition, the positioning error, x_s-x_m, cannot be computed in the controller in ciphertext. The intermediate command is in fact a set of computed multiplication that needs to be decrypted at the plant to compute the final command.

The goal of the proposed research is to establish control theoretic methods to enhance cyber security for networked industrial motion control systems. The proposed approach will encrypt motion control algorithms, feedback and feedforward gains, and sensor signals, and perform necessary computation to generate motion commands to servo systems in the ciphertext space by taking the advantage of homomorphic and somewhat homomorphic encryption. The proposed encrypted motion control configures systems to improve robustness against various types of cyberattacks and prevent sensor data breach and system instability due to controller falsification. The project in particular targets networked motion control systems such as remote assembly systems and robotic manipulators. Research tasks include (1) development of control architecture better suited for homomorphic computation, (2) investigation of quantization errors, time delays and necessary processing resources, (3) implementation of real-time motion control software and hardware with controller encryption. The proposed project will expand the research to homomorphic computation of data-driven gain tuning and low-computational-cost distributed control in the ciphertext space as well as to multilayered security measures with whitelisting.

References:

[1] E. Coumans, Bullet Physics Simulation, in: ACM SIGGRAPH 2015 Courses, ACM, New York, NY, USA, 2015. doi:10.1145/2776880.2792704.

[2] I. Armeni, O. Sener, A.R. Zamir, H. Jiang, I. Brilakis, M. Fischer, S. Savarese, 3D Semantic Parsing of Large-Scale Indoor Spaces, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2016) 1534–1543. doi:10.1109/CVPR.2016.170.

[3] C. Qi, H. Su, K. Mo, L.J. Guibas, PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation, Proc. Computer Vision and Pattern Recognition (CVPR), IEEE. (2017). doi:10.1109/3DV.2016.68.

[4] M. Himmelsbach, F. v. Hundelshausen, H.-. Wuensche, Fast segmentation of 3D point clouds for ground vehicles, in: 2010 IEEE Intelligent Vehicles Symposium, 2010: pages 560–565. doi:10.1109/IVS.2010.5548059.

[5] J. Chen, Y. Fang, Y.K. Cho, C. Kim, Principal Axes Descriptor for Automated Construction-Equipment Classification from Point Clouds, Journal of Computing in Civil Engineering. (2016) 1–12. doi:10.1061/(ASCE)CP.1943-5487.0000628.

[6] K. Glover, D.C. McFarlane, Robust Controller Design Using Normalized Coprime Factor Plant Descriptions, Springer-Verlag, Berlin, Heidelberg, 1989.