Robot-assisted object detection for construction automation : data and information-driven approach

Muhammad Ilyas; Khaw, Hui Ying; Selvaraj, Nithish Muthuchamy; Jin, Yuxin; Zhao, Xinge; Cheah, Chien Chern

2021

Muhammad Ilyas, Khaw, H. Y., Selvaraj, N. M., Jin, Y., Zhao, X. & Cheah, C. C. (2021).
Robot-assisted object detection for construction automation : data and information-driven
approach. IEEE/ASME Transactions On Mechatronics.
<https://dx.doi.org/10.1109/TMECH.2021.3100306>

<https://hdl.handle.net/10356/153478>

<https://doi.org/10.1109/TMECH.2021.3100306>

© 2021 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The published version is available at:
<https://doi.org/10.1109/TMECH.2021.3100306>.

Downloaded on 26 Apr 2023 19:25:11 SGT

Robot-Assisted Object Detection for Construction Automation: Data and Information-Driven Approach

Muhammad Ilyas, Hui Ying Khaw, Nithish Muthuchamy Selvaraj, Yuxin Jin, Xinge Zhao, and Chien Chern Cheah

Abstract—In construction automation, robotics solution is becoming an emerging technology with the advent of artificial intelligence and advancement in mechatronic systems. In construction buildings, regular inspections are carried out to ensure project completion as per approved plans and quality standards. Currently, expert human inspectors are deployed onsite to perform inspection tasks with the naked eye and conventional tools. This process is time-consuming, labor-intensive, dangerous, repetitive, and may yield subjective results. In this paper, we propose a robotic system equipped with perception sensors and intelligent algorithms to help construction supervisors remotely identify the construction materials, detect component installations and defects, and generate report of their status and location information. Building Information Model (BIM) is used for mobile robot navigation and to retrieve building component's location information. Unlike the current deep learning-based object detection which depends heavily on training data, this work proposes a data and information-driven approach which incorporates offline training data, sensor data and BIM information to achieve BIM-based object coverage navigation, BIM-based false detection filtering, and a fine manoeuvre technique to improve on object detections during real-time automated task execution by robots. This allows the user to select building components to be inspected and the mobile robot navigates autonomously to the target components using BIM generated navigation map. An object detector then detects the building components and materials and generates an inspection report. The proposed system is verified through laboratory and onsite experiments.

Index Terms—Robot-Assisted Object Detection (RAOD); Construction inspection; Intelligent monitoring; Building Information Model (BIM); BIM-based navigation; Object coverage.

I. INTRODUCTION

THE construction industry plays a pivotal role in the economic growth of many countries [1], but still, most of the construction works are heavily labor-intensive, dangerous, and inefficient [2]. Therefore, the adoption of robotics and automation has a great potential to revolutionize the industry by providing tremendous improvements in productivity and quality in many ways. From a safety point of view, deploying robotic systems on construction sites greatly reduces possible hazards onsite in conducting dangerous tasks such as manoeuvring heavy and dangerous construction materials and working at high-rise buildings [3]. On top of that, the recent severe outbreak of the COVID-19 pandemic across the world

has upset the construction industry in terms of progress delay, productivity loss, and health hazards as the industry relies heavily on manual processes, which are prone to instrumental and human errors, and fatigue. Therefore, an immense requirement of construction automation has been realized in recent times. In *Automation-in-construction*, many different kinds of robotics or mechatronics systems may be deployed to perform specific tasks. For example, a painting robot [4] can be used to work safely with human co-workers to complete painting tasks. A robot-based steel beam assembly system [5] can be an alternative to ironworkers. Furthermore, in a semi or fully automated construction environment, many systems can interact with each other. The assembly, finishing, and painting tasks in [4], [6], [7] are dependent on the outcome of the progress inspection system, which generates the progress report and instructs other robotic systems to complete the task. A comprehensive description of various construction robots and their applications can be found in [8].

Construction robots can be classified into various categories [8]. We consider the problem of *construction progress monitoring mobile robots* in this work. Inspection is a crucial stage in the progress monitoring of the construction process to ensure work completion in a stipulated time frame adhering to the construction standards. The inspection work can be broadly classified as: *In-Progress Inspection* and *Quality Inspection*. In *In-Progress Inspection*, a supervisor checks the progress with naked-eye and manual instruments whether installation work is completed, and generates a report showing the percentage of work completed at a particular time. The *Quality Inspection* is the final inspection stage carried out before handing over to the customer. This involves rigorous checks of fine details and rectification of defects in the post-construction stage if required. To automate the inspection process, mobile robot systems with various onboard sensors can be used. For example, a post-construction quality inspection robot using scan sensors [9] is used to pick up building defects, such as hollowness, crack, evenness, alignments, and inclination. A mobile robot with a 2D/3D object detection system based on a RGB-D camera is used to update the BIM directly [10]. In [11], a manually-driven mobile robot system with a CCD camera is used for inspecting the cracks in concrete structures using image processing techniques.

This paper aims to develop a mobile robotic system to aid the supervisor in *In-Progress Inspection* in an automated way. The supervisor will load the inspection checklist of a particular floor and the robot navigate autonomously to perform the inspection tasks. The ground mobile robot equipped with intelligent navigation and vision algorithms utilizes multiple sensors

All authors are affiliated with the School of Electrical and Electronic Engineering (EEE), Nanyang Technological University (NTU), 50 Nanyang Ave, Singapore-639798.

This work is supported by the Agency For Science, Technology and Research of Singapore (A*STAR), Singapore, under the National Robotics Program (NRP)-Robotics Domain Specific (RDS: Ref. 1922200001).

Special thanks to Professor Chen I-Ming for providing his support and research resources.

data and BIM information to plan its trajectory throughout the construction area. It can effectively perform tasks like installation check, construction material detection, and construction defects monitoring in a Prefabricated Prefinished Volumetric Construction (PPVC) site and update the inspection checklist automatically. Finally, an updated checklist can be retrieved to prepare a comprehensive inspection report.

To the best of our knowledge, this work is the first to propose a robot-assisted object detection system using multiple forms of data, BIM, and robot sensing information to improve the detection of construction components and materials recognition. Towards achieving this aim, several novel approaches are developed:

a) Object coverage BIM-based navigation: A novel object coverage BIM-based navigation is introduced for the robot-assisted object detection in this paper. Navigation goal points are generated based on the information from the BIM and camera, such that the objects in the inspection checklist are within the field of view (FoV) of the robot vision system. A 2D map and the 3D simulated world are also created utilizing BIM information for navigation.

b) Data and information-driven object detection approach: In addition to a large number of training data collected from the actual PPVC construction site used to train an object detector, the detection model is fed with prior offline information from the BIM model and information from the robot sensors to constraint around the targeted objects in the BIM checklist.

c) BIM-based false detection filtering is developed to validate the detection output with the checklist generated from BIM. This has significantly eliminated false positive detections from the non-targeted objects in the working environment. Furthermore, to inspect small objects from a closer viewpoint, a *fine manoeuvre technique* is used to manoeuvre the robot to a better goal point determined by sensor data, object detection, and BIM information.

The rest of the paper is organized as follows: Section II describes related work. In section III, the technical details of major components of the Robot-Assisted Object Detection (RAOD) system are presented, while the information and data-driven vision system is presented in Section IV. Finally, experimental results and discussions are presented in section V. This paper is concluded in Section VI with some future work insights.

II. RELATED WORK

A. Building Information Modelling and BIM-based navigation

Over the last few decades, the development of BIM has been of great interest in the construction sector. BIM is an integrated process that provides architecture, engineering, and construction (AEC) professionals with tools and technologies to manage the whole life cycle of the building infrastructure. It adopts an object-based parametric modeling technique and generates a 3D digital representation and functional characteristics of the construction site. Many studies have investigated its potential usage in many aspects to assist construction management, such as in the automatic generation of project schedule [12]. In the context of robot navigation and mapping, the semantic and geometric information of building

components and spaces embedded in BIM is exploited for the generation of a navigation maps and design goal points to avoid obstacles in a construction site environment.

Traditionally, architectural drawings are used to solve the navigational problems in mobile robots. For example, Murarka *et al.* [13] used room-level topological map representation for large-scale path planning. Leo Bowen *et al.* [14] provided a comparison of sketched floor plans with SLAM-generated maps and argue that floor plans are closer to the mental maps people would naturally draw to characterize and explore spaces. Behzadian *et al.* [15] tried to solve the navigational issues in environments using hand-drawn sketches but only a coarse localization at room level could be achieved. This raw hand-drawn map approach is not suitable for construction component inspection where accurate self-localization and navigation are required to localize the detected objects in the map in a dynamic environment.

Some researchers have started using BIM-based floor plans for route planning in 2D [16] and 3D [17]. A connection of BIM to Robot operating system (ROS) is explored in [18] to monitor construction progress in which a 4D BIM is used to extract way-points manually. However, an initial navigation map is still created by manually driving a mobile robotic platform. Ratajczak *et al.* [19] proposed a BIM-based augmented reality application for site managers with hand-held devices to visualize the key information related to progress and performance of construction works by using a Location-Based management system with BIM. However, this work presents no robotic solution to the construction automation [20]. Follini *et al.* [21], proposed a system for logistics assistants collaborating with human workers by sharing data between BIM and ROS. A costmap extracted from BIM is used inside the ROS navigation stack. The map is then transmitted and superimposed to the one generated by the LIDAR sensors. However, this work does not fully utilize the BIM information in the generation of a 3D simulated world and in using semantic and geometric information for maximum object coverage.

B. Convolutional Neural Networks Applications in Construction

In recent years, the Deep Convolutional Neural Networks (DCNN) has garnered a lot of attention in providing a promising solution in many diverse areas, such as, in medical science for lung nodule detection [22], diagnosis of mixed faults in rotating machinery [23], traffic-relevant data mining from social media [24] and noise detection and removal [25] in image data. This computer vision-based technology is, however, relatively new in the construction research study. A few discussions have been brought up to address construction-related problems with convolutional neural networks (CNN).

Among them, [26] proved that CNN is a potential perspective in solving construction problems by utilizing feature extraction and classifier to identify asphalt pavement cracks. The CNN model's performance surpassed the traditional image processing technique like the Sobel and Canny edge detector algorithms. A similar study has been carried out by [27] in which a deep Residual Neural Network (ResNet) base network

is used as a backbone for extracting features of the moisture damage to perform detection in asphalt pavement bridges. Both works are significant, but they are not of the scope of monitoring construction conformance. On the other hand, [28] achieved data-free vision-based Faster Region-proposal CNN (Faster R-CNN) to cover four types of construction equipment detection, which are excavator, dump truck, forklift and loader. This work is based on active learning to select the most meaningful and informative data from onsite images to train the deep learning models sequentially using the selected ones. Rashidi *et al.* [29] used classical object recognition methods and, [30] used a pre-trained CNN architecture to identify different construction materials e.g. brick, concrete and wood. Braun *et al.* [31] proposed progress monitoring by acquiring construction images from different viewpoints to produce a 3D point cloud using Structure from Motion (SfM) methods. While CNN is not the fundamental focus, this work proposed the use of Mask Regions with CNN (R-CNN) to further enhance the detection of exterior elements such as columns, walls, formwork, and scaffolding. However, these mentioned works do not target the construction components installation detection, nor pertain to the PPVC environment, as focused in our work. In addition, object detection depends only on the training data and does not take advantage of BIM information of building and sensing information of robot to improve the detection tasks. This paper is, therefore, motivated to develop a RAOD system: an information and data-driven, robot-assisted approach for construction automation to help human supervision in PPVC construction component installations inspection, material recognition, and defect monitoring and localization. The major components of robot-assisted object detection system are explained in the following sections.

III. ROBOT-ASSISTED OBJECT DETECTION

A. Robot-Assisted Object Detection for PPVC Construction Components

The traditional CNN models are inefficient to perform installation and quality inspection in a dynamic environment under varying lighting conditions, such as a construction site, with purely image data. Therefore, to enhance the robustness and reliability of construction inspection applications, a data and information-driven approach for object detection is proposed. On top of the images collected onsite, additional information is extracted from BIM and robot sensors to achieve a reliable object detection. The prior and offline BIM information contains both the semantic and geometric information of the objects to be monitored. The CNN detector is trained with onsite images. However, during the testing time, this approach utilizes both the low and high-level information to perform object detection, which makes the detector more robust and reliable. The robot also uses BIM information along with real-time data from robot sensors, to autonomously navigate to a designated locations to perform the inspection tasks. The proposed construction automation RAOD system is the integration of an autonomous mobile robotic platform with an intelligent object detection system that can be deployed for automated inspection of a construction site. There is a total of

Fig. 1: Dataset for RAOD system can be grouped into four parts: (a) Components installation dataset includes the finished and unfinished windows, doors, and electrical components, (b) materials dataset such as the 3D modules, 2D wall panel, precast staircase, cement, wires, PVC pipes and paint. (c) module gap between module connections, wall defects that can commonly happen during the transportation and hoisting process, and tile defects such as damages, misalignment, and improper joint. (d) Personal Protective Equipment (PPE) of workers onsite.

Fig. 2: Main components of Robot-assisted object detection (RAOD) system for construction inspection: Mobile robot platform; Object detection system on Server, and Control room with BIM software.

six inspection models designed to be executed independently when necessary as shown in Fig. 1. It is designed in a modular and scalable way so that if needed, new models can be trained and easily integrated with the existing system.

B. Robotic System of RAOD

The proposed Robot-assisted object detection system consists of a mobile platform navigating autonomously using the information provided by BIM to perform construction inspection based on the aforementioned detection models. Robot's vision system sends a video stream to cloud server where the data and information-driven detection modules are used for construction component inspection, material identification, and defect detection. The overall RAOD system for construction monitoring system is illustrated in Fig. 2.

A mobile platform (Scout robot) is customized for construction monitoring in this paper. Scout robot is a four-wheel differential drive, skid-steering robot with zero-degree turning radius, driven by 4×200 Watt brushless servo motors. This vehicle weighs about 60-65kg with a payload capacity of 50kg and can move at a maximum speed of 6km/h. It is equipped with a variety of perception and navigation sensors: 3D lidar (RoboSense RS-LiDAR-16), a PTZ IP camera, ultrasonic sensors, GPS, IMU, and wheel encoders. The onboard processing is achieved by Nvidia Jetson AGX Xavier and

Fig. 3: Mechatronics architecture of the mobile robot.

Fig. 4: (a) Generation of ROS map (Occupancy grid map) and 3D simulation world from BIM, (b) generated map and, (c) 3D simulated world.

BeagleBone Black is used as MCU. The Ubuntu 18.04 LTS operating system is running ROS onboard the Jetson computer. The mechatronics architecture of the robot customized for a building inspection is shown in Fig. 3.

C. Object coverage BIM-based navigation for object detection

In RAOD, BIM-based mapping and navigation techniques are introduced. Due to its availability, the BIM is used to create robot navigation maps. For BIM-based navigation, the 2D map and 3D simulated world generation processes are shown in Fig. 4 and described below.

1) *ROS Map Generation from BIM*: In this paper, the BIM model is leveraged to generate 2D navigation map of the environment, without a need of scanning the environment beforehand, as done in conventional approaches to create navigation maps in ROS, e.g. *gmapping*, *hectorSlam*. The 2D floor plan in BIM has rich geometric information of building components such as walls, doors, windows, furniture, etc. The presented approach enables the user to provide the robot with an inspection checklist, then the robot loads the corresponding navigation map and navigates to the desired goal points in the map by avoiding obstacles, thus completely eliminating the need for pre-exploration or mapping of the environment. Furthermore, to run simulations in the generated map, a 3D simulated world of the working environment is required. Conventionally, *Gazebo* simulator is used to draw 3D simulated world, but BIM is employed to generate 3D simulated world in this paper.

2) *Goal Point Design for Object Detection*: In some cases, we need to approach the object to be inspected from a

specific viewpoint. Conventional path planning algorithms do not consider viewing angle and viewing distance to the target object.

In this paper, we consider two aspects of goal point (GP) design: *Viewing distance* to the object, and *viewing angle* of the object. It is assumed that the best viewpoint of a building component is obtained from its front. For example, the coordinates of a door extracted from BIM information is a 2D point at the middle of the bottom of the door. As this point lies beneath the door, a GP has to be designed such that the robot can observe the door from an appropriate distance d covering the object within the camera's field of view (FoV).

To achieve this, the semantic and geometric information of building components is extracted from BIM software. Semantic information includes the component identification number, category of components, and components label, while geometric information contains the dimensions of the components, as well as its location and surface normal vector with respect to the BIM coordinate system. This information provided by BIM is utilized to design the goal points in the working environment.

Let (x_i, y_i, z_i) be the 3D coordinates of the mid-point at the bottom of the i^{th} target object in BIM frame, extracted from BIM information, $F_i = (F_{xi}, F_{yi})^T$ be the surface normal vector that is perpendicular to the i^{th} object surface, then the angle to the F_i with respect to the x -axis of the BIM frame is calculated as:

$$\alpha_i = \tan^{-1}(F_{yi}/F_{xi}) \quad (1)$$

Note that being an indoor mobile robot application, only 2D BIM frame is considered for navigation. The coordinates of the i^{th} goal point (x_{gi}, y_{gi}) for the mobile robot navigation in 2D plane are calculated as:

$$\begin{pmatrix} x_{gi} \\ y_{gi} \end{pmatrix} = \begin{pmatrix} x_{oi} \\ y_{oi} \end{pmatrix} + \begin{pmatrix} \cos \alpha_i & -\sin \alpha_i \\ \sin \alpha_i & \cos \alpha_i \end{pmatrix} \begin{pmatrix} d_i \\ 0 \end{pmatrix} \quad (2)$$

where d_i is calculated from the geometry information, as illustrated in Fig. 5, and $(x_{oi}, y_{oi}) = (x_i, y_i)$ for single object coverage.

Considering an object with total height H_i , i.e. $H_i = z_i + h_i$, where z_i is z-coordinate of the target object, and h_i is the actual height of the object, the width of the object as W_i , camera height d_0 , vertical field of view of camera FoV_v , and horizontal field of view as, FoV_h ; then, the viewing distance d_i is calculated as follows:

$$\begin{aligned} d_{vi} &= \frac{(1+s)(H_{oi} - d_0)}{\tan(\phi_v)} \\ d_{hi} &= \frac{(1+s)(\frac{W_{oi}}{2})}{\tan(\phi_h)} \\ d_i &= \max(d_{vi}, d_{hi}) \end{aligned} \quad (3)$$

where, d_{vi} and d_{hi} are distances between robot and object corresponding to FoV_v and FoV_h respectively, $\phi_v = FoV_v/2$ and $\phi_h = FoV_h/2$, s is a scaling factor set by the user to adjust the percentage of the full object in the image, and $H_{oi} = H_i$, $W_{oi} = W_i$ for coverage of single object. The condition for minimum distance between robot and object is: $d_i = d_{\min}$,

Fig. 5: Goal point generation based on BIM information: An example of, (a) BIM model of a working environment, (b) GP design for two objects, (c) calculation of front distance d based on height of the camera d_0 , object size H , FoV_h and FoV_v , (d) Path planning to GP, (e) object coverage from arbitrary GP, and (f) from BIM-based GP.

when $H_{o_i} \leq 2d_0$. The robot's heading angle, ψ_i , and GP coordinates are calculated as:

$$\begin{aligned} \psi_i &= (\alpha_i - 180) \\ \text{GoalPoint}_i &= (x_{g_i}, y_{g_i}, \psi_i). \end{aligned} \quad (4)$$

However, if two objects are coplanar (i.e. surface normal vectors F are the same), and are close to each other, it is more efficient to generate a single *unified GP* so that camera can cover both objects for detections. For example, if two objects, door and switch board, with BIM coordinates, x_i, y_i , and x_{i+1}, y_{i+1} are close to each other, then the x_{o_i}, y_{o_i} in (2), and W_{o_i}, H_{o_i} in (3) are calculated as follows:

$$\begin{aligned} x_{o_i}, y_{o_i} &= \frac{x_i + x_{i+1}}{2}, \frac{y_i + y_{i+1}}{2} \\ H_{o_i} &= \max(H_i, H_{i+1}) \\ W_{o_i} &= \frac{W_i}{2} + B_i + \frac{W_{i+1}}{2} \end{aligned} \quad (5)$$

where $H_i, H_{i+1}, W_i, W_{i+1}$ are the height and width of the two consecutive objects in the BIM checklist respectively, and B_i is the distance between the two object's centers. Objects under inspection are fully covered when the robot is placed at BIM-based designed GP, as contrary to selecting arbitrary goal point using *rviz* utility, as shown in Fig. 5 (e) and (f).

The goal point (x_{g_i}, y_{g_i}) is then passed to a ROS package, called *move_base*, which will attempt to reach the goal point in the BIM frame, with a mobile base. The ROS path planner module, e.g. Dynamic Window Approach (DWA) is typically used as a local path planner and A^* , as a global path planner. Algorithm-1 takes BIM information as input and performs navigation tasks to reach the GP of an object under inspection. It then generates a *Flag* signal to the vision system to starts

the detection task when the robot reaches the designated goal point. If detection is successfully completed, this algorithm then checks the next two objects in the BIM checklist to generate the next navigation goal point.

Algorithm 1: BIM based navigation to cover objects in view to perform detection tasks. *Flag₁* is sent from the navigation system to the vision system to start the detection task, and *Flag₂* is sent back from the vision system to navigation. Distance threshold, T_r between two objects is chosen as 1.5m

Input: BIM information (Semantic and geometric info)

Output: *Flag₂*

```

1 initialization;
2  $Flag_1 = 0, Flag_2 = 0$ ;
3  $N \leftarrow BIMchecklist$ ;
4 while  $i \leq (N - 1)$  do
5   Check distance ( $B$ ) between objects  $(x_i, y_i)$  and
    $(x_{i+1}, y_{i+1})$ , and corresponding surface angles,
    $\alpha_i$  and  $\alpha_{i+1}$ ;
6   if  $H_{o_i} \leq 2d_0$  then
7      $H_{o_i} = 2d_0$ 
8   end
9   if ( $B < T_r$  and  $\alpha_i = \alpha_{i+1}$ ) then
10    Equation(2-5);
11  else
12    Exclude  $(i + 1)^{th}$  object for GP calculation;
13    Equation(2-4);
14  end
15  ROS Move Base Package  $\leftarrow GoalPoint_i$ ;
16  if (GP reached) then
17     $Flag_1 = 1$ ;
18    Send  $Flag_1 = 1$  to vision system to start
    detection task.
19  end
20  if (Check  $Flag_2 = 1$ ) then
21    if  $Flag_2 = 1$ , then choose next two Objects in
    BIM checklist for GP calculation.
22  end
23 end

```

IV. CONSTRUCTION COMPONENTS INSTALLATION AND DEFECT DETECTION BASED ON DATA AND INFORMATION-DRIVEN CNN DETECTOR: VISION SYSTEM

A. YOLO as the Backbone of Detection Algorithm for Components Installation Check

You Only Look Once (YOLO) [32] is a CNN-based single stage object detector that processes an image in a single framework by performing detection as a regression problem to predict bounding box coordinates and their associated class probabilities. It has low inference time due to its simple and efficient architecture, making it the best choice for faster real-time object detection applications today. Among the YOLO family, YOLOv3 is the state-of-the-art detector with a 53-layer CNN feature extractor, and this deep architecture is the backbone of the data and information-driven CNN detector

Fig. 6: The overall data and information based CNN detector for construction components installation and defect detection.

proposed in this paper. To train the detection models, more than 5000 training images were collected from an ongoing PPVC construction project site. The training images were collected over months to cover the PPVC construction sites at different stages of the project completion. These images were taken under varying lighting conditions and at different angles and distances in multiple sites to get a diverse training dataset. To ensure the desired detection performance, all images were captured in good quality and in focus. The YOLOv3 utilizes residual connections [33] and performs detection across three different scales, like, the feature pyramid networks (FPN) [34]. The input image is fed into the CNN feature extractor, and subsequent up-sampling along with concatenation of previous layers results in a feature map of three different scales. Each scale corresponds to a $S \times S$ grid in the input image where each grid cell predicts B-bounding boxes with x and y coordinates (b_x, b_y), and bounding box width and height (b_w, b_h) using linear regression, its objectness score O using logistic regression and the corresponding class of the object within the bounding box using binary cross-entropy. Each grid is assigned a set of anchor boxes with dimensions (p_w, p_h), and YOLO predicts the bounding box dimensions relative to the anchor boxes, which reduces the range of values of predictions. Finally, the detections of all three layers undergo Non-Maximal Suppression (NMS) to eliminate multiple overlapping predictions. However, the YOLOv3 detector suffers from high variance in bounding box predictions, false detection, and misclassifications. Thus, the moving average is applied to avoid the coordinates of the bounding boxes from changing dramatically with respect to time. On top of that, to ensure detection consistency, detections across multiple frames are processed using a K-means clustering algorithm which outputs only the detections that are consistent across multiple frames and removes sparsely occurring false detections. To alleviate false detections and misclassifications, the visual information alone is insufficient. Hence, a data and information-driven approach is proposed in this work, where metadata from BIM and onboard sensors information are utilized to remove false detections and to ensure reliable components installation and defect detection in the PPVC construction site, as shown in Fig. 6.

Fig. 7: (a) Camera frame and BIM frame, (b) Object detection in image plane

B. BIM-based False Detection Filtering

Some construction components share similar visual features, especially when captured from a certain angle under particular illumination conditions causing inconsistent detection. This may undesirably give rise to false detection when the K-means algorithm is not able to eliminate the consistent detection. To solve this problem, the geometrical information from BIM, such as the object's location, orientation, dimensions are used to filter out false detections.

Object localization is required before BIM-based filtering. Firstly, the camera is calibrated to get intrinsic parameters, e.g. image center point (c_x, c_y), focal length (f), both in pixels, and distortion coefficients, as illustrated in Fig. 7. These parameters are used to rectify the image and localize the detected object.

From the YOLO detection results, the centers of the objects in the image frame, b_x and b_y are known. As the robot is facing the target object for inspection, in this application, the distance D to the detected object is directly taken from a frontal ray of the lidar sensor, and the 2D image points are converted to 3D points in camera frame as follows:

$$\begin{pmatrix} x_{cam} \\ y_{cam} \\ z_{cam} \end{pmatrix} = \begin{pmatrix} (b_x - c_x) * D / f \\ (b_y - c_y) * D / f \\ D \end{pmatrix} \quad (6)$$

Furthermore, to get object location in BIM frame, 3D camera frame points are transformed to BIM frame through successive homogeneous transformations, as follows:

$$\mathbf{X}^{BIM} = T_{map}^{BIM} \cdot T_{robot}^{map} \cdot T_{lidar}^{robot} \cdot T_{cam}^{lidar} \cdot \mathbf{X}^{cam} \quad (7)$$

where T_a^b represents the homogeneous transformation matrix from frame-a to frame-b, $\mathbf{X}^{BIM}$, $\mathbf{X}^{cam} \in \mathbb{R}^{4 \times 1}$ are position vectors in homogeneous coordinate of BIM frame and camera frame, respectively.

The localized object's size, location, and orientation are then compared with BIM information. The detected components that do not agree with the BIM ground truth, are removed automatically from the detection list. The filtered detection output is, therefore, a bilateral verification from both the prior offline BIM information and the real-time detection output from the detector. With BIM-based filtering, the data and information-driven detection model provides a robust and intelligent automated solution to installation and defect inspection.

C. Fine manoeuvre using Sensor Information

When the robot executes BIM-based navigation as given in Algorithm-1, the final GP may not be reachable in some cases due to inaccuracies in the navigation system. In cases where small objects are to be observed from a closer viewpoint, a new goal point can be determined by the fine manoeuvre technique. The detections obtained from the current goal point can be used as visual information for the robot to perform fine manoeuvres to reach a better viewpoint to get detections with higher confidence. Two values namely, *yaw* and *move distance*, (YM) are calculated [35] from the current detections and the camera's intrinsic parameters for fine-grain rotation and linear adjustment to better observe the object under inspection. The *yaw* angle is calculated from the difference between the image and bounding box centers as,

$$yaw : \Delta\theta_{xi} = \tan^{-1}\left(\frac{c_x - b_{xi}}{f}\right) \quad (8)$$

where positive and negative values of *yaw*($\Delta\theta_{xi}$) angle indicate rotation in the clockwise and anti-clockwise direction, respectively.

The *move distance* is calculated by making fine adjustments to the current distance D between the robot and the object. Since the detection output is relative to the anchor boxes, therefore by achieving a viewpoint with object dimensions in an image frame resembling that of anchor box dimensions will ensure high confidence detections. The magnitude of change is calculated from the difference between the current object height and the anchor box height of that object, as follows,

$$move\ distance : M_{di} = D(1 - \frac{b_{hi}}{p_{hi}}) \quad (9)$$

where, b_{hi} is the object height of the i^{th} object in image plane obtained from the detection and p_{hi} is the anchor box height of the i^{th} object. Positive M_d value indicates a forward movement of the robot towards the object while a negative value refers to backward motion.

Since the camera is fixed and the robot cannot execute pitch movements, there are chances that the objects can go outside the Field Of View (FOV) while performing the fine manoeuvre. Hence, an upper limit on the move distance, M_d^{max} , is formulated which ensures that objects do not go outside the FOV when moving forward, as follows,

$$Max.\ Move\ distance : M_d^{max} = D(1 - \frac{\beta}{I_h/2}) \quad (10)$$

$$where, \beta = \max(|b_{yi} - c_y| + (b_{hi}/2)) ; \forall i$$

For multiple objects in a single frame, the Y, M values are calculated individually for every object and the arithmetic mean of yaw values and maximum value of move distances $\max(M_{di})$ is considered as the final output. If the calculated move distance is beyond the upper limit, then the maximum move distance M_d^{max} is considered as the final output. The calculated *yaw* and *move distance* values are with respect to the camera frame, and to move the mobile robot to the

fine goal point, these values are first converted into displacement in the 2D x and y axis as follows,

$$\begin{aligned} \Delta x_i &= M_{di} \cdot \sin \Delta\theta_{xi} \\ \Delta y_i &= M_{di} \cdot \cos \Delta\theta_{xi} \end{aligned} \quad (11)$$

and then converted to robot frame, using frame transformations defined in Eq. 7. The displacement in x and y direction with respect to the robot frame, $\Delta x_i^r, \Delta y_i^r$, are converted to fine goal point in BIM frame as follows:

$$\begin{pmatrix} x_{gi}^F \\ y_{gi}^F \end{pmatrix} = \begin{pmatrix} x_{gi} \\ y_{gi} \end{pmatrix} + \begin{pmatrix} \cos \psi_i & -\sin \psi_i \\ \sin \psi_i & \cos \psi_i \end{pmatrix} \begin{pmatrix} \Delta x_i^r \\ \Delta y_i^r \end{pmatrix} \quad (12)$$

where x_{gi}^F, y_{gi}^F is the *fine manoeuvre GP* in 2D, x_{gi}, y_{gi} is current position, and ψ_i is the current heading of the robot. The heading of the robot at fine GP is: $\psi_i^F = (\alpha_i - 180)$.

V. EXPERIMENTAL SETUP, RESULTS AND DISCUSSION

This section presents and discusses the detection results of data collected from construction site and supplementary testing done in labs to verify the proposed data and information-driven Robot-Assisted Object Detection approach for construction automation. The images for training YOLOv3 model were collected in an ongoing PPVC construction site for residential flats by using a handheld camera.

From the training dataset, six detection models, namely, the Components installation model, Materials check model, PPVC module gap inspection model, Wall defect check model, Tile defect check model, and Workers PPE inspection model are trained separately on a $4 \times$ Tesla V100 GPU server running on Linux platform. During the training phase, the models are trained at a learning rate of 0.001, 4000 steps per class, and a momentum of 0.9 on input images of size 608 x 608. Data augmentation is used to improve model generalization. The training time needed to train a model is approximately 8 hours. The testing images are from a separate dataset and only the detections with a confidence level higher than a threshold of 0.8 are considered. The detection result from the trained object detector is combined with the information extracted from BIM to realize the proposed data and information-driven object detector.

The experimental results are organized into two parts: First, detection results based on testing videos and images obtained from the actual construction site by using the handheld camera are presented. Second, experimental results based on the proposed Robot-assisted object detection system are presented. However, due to the activity restrictions enforced on actual construction sites during the pandemic, the experiments using the mobile robot system were only performed in the lab environment.

A. Detection Results from Construction Site Dataset

Detection models are trained and tested on this dataset to perform detection tasks, such as installation check, construction material detection, and construction defect detection on PPVC building components.

Fig. 8: Detection of components and materials on actual site images: (a) detection of door, window and electrical components, (b) detection of uninstalled door, electric panel and staircase, (c) detection of building materials: 2D wall, cement bag and PPVC block.

1) *Components installation monitoring:* The performance of the YOLOv3 detector model is evaluated on components installation check and materials detection in terms of mean Average Precision (mAP). From the training data, a baseline mAP of 74.13% is achieved for the installation check model and 47.57% for the materials detection model. The mAP for installation components is calculated from 220 test images with an Intersection over Union (IoU) threshold at 0.8 and 0.5 for construction materials with 152 images. The detection results of building components at the actual construction site are given in Fig.8. Detection of installed building components is given (a) and (b), and the detection of building material and PPVC blocks is shown in (c). The results show that the trained YOLOv3 model is able to detect building components and materials and report their status accordingly.

2) *Installation gap inspection and defect check:* The detection results for various defects in the construction process are presented in this section. The detection models for PPVC module gap inspection, wall crack check, and tile defect check are independent and trained separately. During the test time, the individually trained models are grouped together to form a defect check module capable of performing multiple defect check tasks simultaneously. Fig. 9 (a) and (b) show the module gap detection results while tiles defects, tiles misalignment and wall cracks are presented in (c), (d) and (e) respectively.

B. RAOD Detection Results in Lab Environment

In this section, we present RAOD detection results performed in a lab environment using the mobile robot system as shown in Fig. 2 and Fig. 5. The YOLOv3 model trained on real site images is tested on lab data. The localization of different building components and defects is performed.

1) *Localization of building materials and wall cracks:* As the BIM model has no construction materials and defects information, their location can not be compared with BIM data and pre-calculated goal points can not be designed to detect

Fig. 9: Defects check module: detection results for (a) unfilled gap between staircase module and corridor, (b) filled gap between two PPVC blocks, (c) misalignment between tiles, (d) tile damages, and (e) wall crack detection by RAOD in real PPVC data set.

Fig. 10: Material and wall cracks localization: (a) 2D detection in camera frame, with a closeup view of crack on left, (b) 3D localization in lidar frame.

building material. Therefore, the *sensorfusion* approach is utilized to localize building materials and wall defects. To obtain 3D location information in the BIM frame, the detection bounding boxes of building material in the camera frame are transformed in the BIM frame with the help of camera intrinsic parameters and a transformation matrix between camera and lidar sensors as given in equation (7). Note that BIM-frame origin and ROS map origin are matched before starting the experiments. Fig. 10 shows the detection and localization results of some of the common building materials e.g. cement bags, electrical wires in a lab environment, and localization results of these materials are presented in Table I.

Similarly, to check the wall cracks, the user selects a wall to be inspected and the robot autonomously navigates to the targeted wall and scans it for any defects. The robot scans the wall with camera and lidar sensors, and with the help of sensor fusion, localizes wall cracks and defects in the BIM frame. During these experiments, the current date and time are also recorded to keep track of the inspection time.

The absolute positions of the materials and wall cracks were measured manually with a measuring tape as Ground truth, with respect to BIM-origin in the robot navigation map, as it was inexpensive and the measurements were sufficiently accurate for our purpose.

The localization error results with respect to BIM coordinates for each detected wall crack are shown in Table I. Note that the average localization error for material detection and wall cracks is $\leq 5.3\text{cm}$ in this experiment, which is sufficient to locate the detected object and notify the site supervisor.

TABLE I: Construction material and Wall crack localization. Electrical wires and cement bags are some common building materials onsite. *G. Truth* is ground truth obtained from tape measure. Two cracks were detected during scanning this wall in lab experiment. The average localization error is given as (*Av.E*) in cm.

Name	-	x (cm)	y (cm)	z (cm)	Date/Time
Cement bag	G.Truth	-305.0	-220	45	24.12.20
	RAOD	-307.3	-229.0	40.5	16 : 29 : 30
Error(cm)	-	2.3	9.0	4.5	Av.E=5.3
Elect. wires	G.Truth	-305.0	-305.0	60	24.12.20
	RAOD	-303.1	-312.9	54	16 : 29 : 30
Error(cm)	-	1.9	7.9	6	Av.E=5.2
Crack1	G.Truth	65	460	80	24.12.20
	RAOD	60.9	455.1	83.5	17 : 35 : 50
Error(cm)	-	4.1	4.9	3.5	Av.E=4.1
Crack2	G.Truth	155	460	80	24.12.20
	RAOD	159	458.3	84.1	17 : 55 : 10
Error(cm)	-	4.0	1.7	4.1	Av.E=3.3

2) *BIM information-based false detection filtering and building components localization:* Although achieving an mAP of 74.13% onsite images, the YOLOv3 detector is prone to false detection. The proposed data and information-driven approach make use of prior meta information to perform false detection filtering by first localizing the detected bounding boxes with the mobile robot's location and transformation matrix information. Then, the ground truth of the object's location is extracted from the BIM information. The component location, dimensions, surface normal are used to calculate the overlapping of localized values with ground truth in terms of IoU, and the components having IoU lower than a threshold are filtered out.

An example of how the proposed detector filters out false detection using BIM information is shown in Fig. 11.

Six electrical switches are detected including a false detection on a thermal controller while (b) shows that the proposed approach has accurate detection with BIM-based filtering. Table II shows that the falsely detected object (i.e. *Switch 6*) has zero IoU while others have IoU larger than 0.5. Thus, false

TABLE II: BIM information-based filtering by using object location in BIM, and calculating IoU to filter out false detections. *G.Truth* is ground truth location of objects in the BIM frame.

Objects in Checklist	-	Location (x, y, z (cm))			IoU	Status
Switch1	G.Truth	-200.25	0	142	0.53	Installed
	ROAD	-198.89	0	143.10		
Switch2	G.Truth	-196.25	0	130	0.55	Installed
	ROAD	-195.59	0	131.99		
Switch3	G.Truth	-206.75	0	130	0.56	Installed
	ROAD	-205.68	0	131.99		
Switch4	G.Truth	-218.25	0	130	0.60	Installed
	ROAD	-216.94	0	131.97		
Switch5	G.Truth	-228.75	0	130	0.62	Installed
	ROAD	-227.81	0	132.38		
Switch6	G.Truth	-	-	-	0	False Detection
	ROAD	-216.94	0	147.39

Fig. 11: False detection filtering: (a) YOLOv3 detector detects six switches and one of them is a false detection, (b) Data and Information-driven detector filters out the false detection by comparing with BIM information.

Fig. 12: (a) Initial GP generated to detect both door and switches, (b) Fine manoeuvre performed for a closer view of switches.

detection is filtered out with BIM-based filtering technique in the proposed RAOD system. The experimental results of the proposed data and information-driven approach have shown that the heavy dependency of an object detector on the training data can be alleviated with the utilization of BIM information. This approach is not restricted to YOLOv3 but can be extended to any other object detector, thereby improving its performance.

3) *Vision-based fine manoeuvre of the mobile robot:* The experimental results of the mobile robot's fine manoeuvre are presented in Fig. 12. From the BIM checklist, an initial goal point is generated to check the door and the switches at the same time. However, since the switches are small objects, they are not clearly visible from the current position.

To improve detection performance, a fine manoeuvre is performed to achieve a better and closer view of the switches. From the initial viewpoint, a yaw value of 2.26° and a move distance of 102cm are calculated for the switches. The threshold values for the yaw and move distance are set to be 2° and 5cm, respectively to ensure high precision manoeuvres. The calculated values are in the camera frame and these are converted into GP in the BIM frame and passed on to the

mobile robot's navigation system to perform fine manoeuvre. The mobile robot takes these values as input and executes fine manoeuvre by moving to the target GP, thus achieving the desired viewpoint. The switches are perfectly centered in the image with a clear and closer view while the yaw and move distance values are within the threshold after the fine manoeuvre is executed.

4) *Safety inspection module*: The safety module is to detect multiple safety equipment of workers, such as face masks, safety helmets, boots, and vests, as shown in Fig. 13. From the bounding box centers of the different PPE and their relative positions with the person's bounding box center, the conclusion is drawn if a worker is wearing full PPE.

C. Comparison of RAOD detector and YOLOv3

An experiment was performed to compare the performance of the proposed RAOD system with conventional YOLOv3. BIM-based designed goal points were used to obtain enhanced detections to initiate the detection task.

TABLE III: Comparison of RAOD detector with conventional YOLOv3 for building components detection. *F.D* and *M.D* are number of False Detections and Misdetctions, respectively.

Obj	Metric	YOLOv3				RAOD	
		Conf. (%)	F.D	Cov. (%)	M.D.	Conf. (%)	Cov. (%)
Main Door	Mean	72.6	7	28.2	0	98.9	48.8
	Std. Dev	21.2	-	19.4	-	2.8	5.4
Back Door	Mean	91.6	3	32.2	0	98.9	47.4
	Std. Dev	12.6	-	14.1	-	1.2	5.1
Switch Board	Mean	90.1	3	6.2	3	97.3	8.3
	Std. Dev	5.3	-	4.1	-	2.1	0.8

In this experiment, 10 different users were asked to command the robot randomly at human-judged goal points, to inspect three building components: Main door, back door, and switchboard. For each experiment, detection confidence, percentage of object coverage, and data standard deviation were recorded, and results were summarized in Table. III. The object coverage is calculated as: $\sum_{i=1}^n \frac{A_D^i}{A_T}$, where A_D and A_T are detected area and the total area of the image respectively, and n is the number of detected objects. The average detection confidence and object coverage was high with low variance in the RAOD detector. With BIM-based filtering, there was no observed false detection in RAOD surpassing the conventional YOLOv3 detector performance in which seven false detections were observed in the main door, three in the back door, and three in detecting switches. On top of that, misdetections also occurred three times in the YOLOv3 detector when detecting small objects like switches, but RAOD showed no misdetection in all the experiments. The proposed RAOD detector, therefore, outperforms the conventional approach in terms of detection confidence, object coverage, numbers of false detections, and the number of misdetections.

VI. CONCLUSION AND FUTURE WORK

A Robot-Assisted Object Detection (RAOD) system has been proposed in this paper. This system integrates the mobile robotic platform, BIM information, and image data-driven YOLOv3 detector to perform the object detection tasks in a PPVC construction site in pursuit of construction automation.

Fig. 13: PPE safety monitoring: (a) Detection of Personal Protective Equipment (PPE) and inference by the safety inspection module, (b) Localization of workers.

BIM-based map generation is introduced for robot navigation to the specific goal points designed for maximum object coverage using BIM information and sensor data. Experimental results demonstrate that incorporating online and offline data with information from BIM alleviates the problems inherent to conventional detectors, such as false detection filtering, which are minimized by leveraging BIM information. Furthermore, object coverage is maximized and misdetections are minimized with the use of proposed *fine manoeuvre technique*, thus reinforcing the efficacy of utilizing the data and information-driven approach in object detector (RAOD). The proposed system is, thereby, readily available to perform several functions in construction automation; for instance, components installation detection, building materials check and localization, module gap inspection, building component defects check, and worker's safety PPE detection. This work can be further extended to multi-floor navigation in a PPVC construction site by installing a communication system between the elevators and the robot, enabling the wheeled robot to navigate between different levels within the same building. To move between indoor and outdoor environments, the robot needs to switch from GPS sensing to the BIM-based indoor navigation system. However, in a non-PPVC site where elevators are not yet ready, it is necessary to use a legged robot or drone for the inspection tasks.

Due to pandemic restrictions of onsite visits, validation of the proposed robot system has been done through experiments conducted in the lab environment. As the situation improves, actual onsite tests are planned to validate multi-unit multi-room autonomous navigation using BIM model and achieve an automated system that can be deployed onsite to help in daily construction inspection tasks in PPVC sites.

REFERENCES

- [1] T. D. Oesterreich and F. Teuteberg, "Understanding the implications of digitisation and automation in the context of industry 4.0: A triangulation approach and elements of a research agenda for the construction industry," *Computers in industry*, vol. 83, pp. 121–139, 2016.
- [2] R. Kangari and D. W. Halpin, "Potential robotics utilization in construction," *Journal of Construction Engineering and Management*, vol. 115, no. 1, pp. 126–143, 1989.
- [3] J. M. D. Delgado, L. Oyedele, A. Ajayi, L. Akanbi, O. Akinade, M. Bilal, and H. Owolabi, "Robotics and automated systems in construction: Understanding industry-specific challenges for adoption," *Journal of Building Engineering*, vol. 26, p. 100868, 2019.
- [4] E. Asadi, B. Li, and I.-M. Chen, "Pictobot: a cooperative painting robot for interior finishing of industrial developments," *IEEE Robotics & Automation Magazine*, vol. 25, no. 2, pp. 82–94, 2018.
- [5] B. Chu, K. Jung, M.-T. Lim, and D. Hong, "Robot-based construction automation: An application to steel beam assembly (part i)," *Automation in Construction*, vol. 32, pp. 46–61, 2013.

- [6] I. D. Spath, I. J. Schmidt, and D. Andres, "Robot systems for residential building construction: integration and operation," *Proceedings of the 14th ISARC*, pp. 185–197, 1997.
- [7] K. Iturralde, M. Feucht, R. Hu, W. Pan, M. Schlandt, T. Linner, T. Bock, J.-B. Izard, I. Eskudero, M. Rodriguez, *et al.*, "A cable driven parallel robot with a modular end effector for the installation of curtain wall modules," in *Proceedings of the 37th International Symposium on Automation and Robotics in Construction (ISARC)*, Kitakyushu, Japan, 2020, pp. 27–28.
- [8] T. Bock and T. Linner, *Construction Robots: Volume 3: Elementary Technologies and Single-task Construction Robots*. Cambridge University Press, 2016.
- [9] R.-J. Yan, E. Kayacan, I.-M. Chen, L. K. Tiong, and J. Wu, "Quicabot: Quality inspection and assessment robot," *IEEE Transactions on Automation Science and Engineering*, vol. 16, no. 2, pp. 506–517, 2018.
- [10] M. Ferguson and K. Law, "A 2d-3d object detection system for updating building information models with mobile robots," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2019, pp. 1357–1365.
- [11] S.-N. Yu, J.-H. Jang, and C.-S. Han, "Auto inspection system using a mobile robot for detecting concrete cracks in a tunnel," *Automation in Construction*, vol. 16, no. 3, pp. 255–261, 2007.
- [12] H. Kim, K. Anderson, S. Lee, and J. Hildreth, "Generating construction schedules through automatic data extraction using open bim (building information modeling) technology," *Automation in Construction*, vol. 35, pp. 285–295, 2013.
- [13] A. Murarka and B. Kuipers, "Using cad drawings for robot navigation," in *2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace (Cat. No. 01CH37236)*, vol. 2. IEEE, 2001, pp. 678–683.
- [14] L. Bowen-Biggs, S. Dazo, Y. Zhang, A. Hubers, M. Rueben, R. Sowell, W. D. Smart, and C. Grimm, "Sketched floor plans versus slam maps: A comparison," *arXiv preprint arXiv:1606.04836*, 2016.
- [15] B. Behzadian, P. Agarwal, W. Burgard, and G. D. Tipaldi, "Monte carlo localization in hand-drawn maps," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2015, pp. 4291–4296.
- [16] C. Schlette and J. Roßmann, "Sampling-based floor plan analysis on bims," in *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, vol. 33. IAARC Publications, 2016, p. 1.
- [17] W. Lin, P. Lin, and H. Tserng, "Automating the generation of indoor space topology for 3d route planning using bim and 3d-gis techniques," in *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, vol. 34. IAARC Publications, 2017.
- [18] A. Ibrahim, A. Sabet, and M. Golparvar-Fard, "Bim-driven mission planning and navigation for automatic indoor construction progress detection using robotic ground platform," 2019.
- [19] J. Ratajczak, C. Marcher, C. P. Schimanski, A. Schweikopfler, M. Riedl, and D. T. Matt, "Bim-based augmented reality tool for the monitoring of construction performance and progress," Chania, Proceedings of the 2019 European Conference for Computing in ..., 2019.
- [20] Y. Chen, S. Huang, and R. Fitch, "Active slam for mobile robots with area coverage and obstacle avoidance," *IEEE/ASME Transactions on Mechatronics*, 2020.
- [21] C. Follini, M. Terzer, C. Marcher, A. Giusti, and D. T. Matt, "Combining the robot operating system with building information modeling for robotic applications in construction logistics," in *International Conference on Robotics in Alpe-Adria Danube Region*. Springer, 2020, pp. 245–253.
- [22] J. Zheng, J. Guo, X. Cui, R. N. Veldhuis, M. Oudkerk, and P. M. Van Ooijen, "Automatic pulmonary nodule detection in ct scans using convolutional neural networks based on maximum intensity projection," *IEEE transactions on medical imaging*, vol. 39, no. 3, pp. 797–805, 2019.
- [23] S. Chen, Y. Meng, H. Tang, Y. Tian, N. He, and C. Shao, "Robust deep learning-based diagnosis of mixed faults in rotating machinery," *IEEE/ASME Transactions on Mechatronics*, vol. 25, no. 5, pp. 2167–2176, 2020.
- [24] Y. Chen, Y. Lv, X. Wang, L. Li, and F.-Y. Wang, "Detecting traffic information from social media texts with deep learning approaches," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 8, pp. 3049–3058, 2018.
- [25] H. Y. Khaw, F. C. Soon, J. H. Chuah, and C.-O. Chow, "High-density impulse noise detection and removal using deep convolutional neural network with particle swarm optimisation," *IET Image Processing*, vol. 13, no. 2, pp. 365–374, 2018.
- [26] H. Nhat-Duc, Q.-L. Nguyen, and V.-D. Tran, "Automatic recognition of asphalt pavement cracks using metaheuristic optimized edge detection algorithms and convolution neural network," *Automation in Construction*, vol. 94, pp. 203–213, 2018.
- [27] J. Zhang, X. Yang, W. Li, S. Zhang, and Y. Jia, "Automatic detection of moisture damages in asphalt pavements from gpr data with deep cnn and irs method," *Automation in Construction*, vol. 113, p. 103119, 2020.
- [28] J. Kim, J. Hwang, S. Chi, and J. Seo, "Towards database-free vision-based monitoring on construction sites: A deep active learning approach," *Automation in Construction*, vol. 120, p. 103376, 2020.
- [29] A. Rashidi, M. H. Sigari, M. Maghiar, and D. Citrin, "An analogy between various machine-learning techniques for detecting construction materials in digital images," *KSCE Journal of Civil Engineering*, vol. 20, no. 4, pp. 1178–1188, 2016.
- [30] S. Bunrit, N. Kerdprasop, and K. Kerdprasop, "Evaluating on the transfer learning of cnn architectures to a construction material image classification task," *Int. J. Mach. Learn. Comput.*, vol. 9, no. 2, pp. 201–207, 2019.
- [31] A. Braun, S. Tüttas, A. Borrmann, and U. Stilla, "Improving progress monitoring by fusing point clouds, semantic data and computer vision," *Automation in Construction*, vol. 116, p. 103210, 2020.
- [32] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.
- [33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
- [34] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117–2125.
- [35] N. M. Selvaraj, I. Muhammad, and C. C. Cheah, "Towards dependable object detection," in *IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society*, 2020, pp. 523–528.

Muhammad Ilyas received Ph.D. in Robotics and Virtual Engineering from the UST, S. Korea in 2016. He is currently a Research Fellow at Nanaynag Technological University (NTU), Singapore. His area of research is Robot navigation and application of Deep Learning, RL, in robotics.

Hui Ying Khaw completed her Ph.D. degree from University of Malaya, Malaysia, in 2019. She was a postdoctoral research fellow in School of EEE, NTU and is now with Monash University Malaysia as a lecturer. Her research areas are object detection and deep learning.

Nitish, M. Selvaraj received B.E. in Electronics and Communication Engineering from Anna University, India, and M.Sc from School of EEE, NTU, Singapore. He is currently a RA in the School of EEE, NTU. His research interests include computer vision and deep learning.

Jin Yuxin received her B.E. in computer engineering from SUTD, Singapore, in 2019. She is currently pursuing her M.Eng degree in School of EEE, NTU, Singapore. Her research interests include control theory, robot vision, and deep learning based robotic control.

Zhao Xinge received her B.Eng. in Civil Engineering, from the NTU, Singapore in 2019. She is currently pursuing her M.Eng degree in School of EEE, NTU, Singapore. Her research interests lie in the area of autonomous robot and automation in construction.

Chien Chern Cheah received B.E. degree in electrical engineering from Nanyang Technological University (NTU), Singapore in 1990, and the M.E. and Ph.D. degrees in electrical engineering, from NTU, Singapore, in 1993 and 1996, respectively. He is an Associate Professor with the School of Electrical and Electronic Engineering, NTU. He is an Associate Editor for Automatica.