Geometrical-Based Displacement Measurement With Pseudostereo Monocular Camera on Bidirectional Cascaded Linear Actuator

Denzel Lee, Jingmin Liu, Ryan Lim, Jie Lin Chan, and Shaohui Foong

Abstract—This article details the development of a geometrical-based displacement extraction framework capable of automatically extracting critical infrastructure measurements in one sequence. The framework is a novel rail viaduct bearing inspection pipeline implemented on Bearing Inspector for Narrow-space Observation Version 2 (BINOv2). BINOv2 is a tethered custom unmanned aerial vehicle system utilized to supplant labor-intensive pipelines and enhance inspection accuracy of infrastructure conditions in confined remote locations. The algorithm accepts stereoscopic images taken from a single monocular camera on a bidirectional cascaded linear actuator system in a rack-and-pinion configuration. A point cloud model generated from the image sets then runs through a hierarchical neural network for 3-D segmentation to extract targeted regions of interest. Our training pipeline generates and forms the full model's training dataset using only a small sample of real point clouds. The point cloud generated is inadequate to form the full bearing geometry profile. Therefore, the proposed framework projects best-fit circles based on the point cloud curvature to form the full bearing geometry profile so that the required displacement measurement is available for extraction. Several experiments were conducted on a mock-up and actual operational site to validate the proposed framework's accuracy, its robustness and comparison with other state-of-the-art alternatives.

Index Terms—Actuators, condition monitoring, displacement measurement, learning (artificial intelligence), 3-D segmentation, unmanned aerial vehicles (UAVs).

I. INTRODUCTION

METROLOGY is an integral part of structural inspection. However, some inspection targets' location is remote or inaccessible, rendering it challenging to collect critical measurements; for example, powerlines, wind turbines, civil infrastructure, to name a few. This challenge makes it inherently dangerous, cumbersome, expensive, and inefficient for the inspection process. Thanks to the exponential growth in robotics, computer science, and many other related fields, drone or unmanned aerial vehicle (UAV) has served as an optimal solution when coupled with computer vision methods [1], [2]. This article is an extension of our previous work [3], where the application in focus is the inspection of the rail viaduct bearing (see Fig. 1). The bearing, which houses inside a cavity, structurally connects the separately constructed pier and viaducts. The current inspection method involves a civil engineer visually inspect a scale bar the bearing from the outside of the cavity, usually at the height of 5 m to 10 m. However, most bearing scale bars readings have faded due to wear and tear. Hence, it is unusable for the inspection process. Our proposed framework is not reliant on the readings.

The maximum operational space available within the cavity is 850 mm in length, 150 mm in width, and 300 mm in depth. Some key features of the proposed algorithm to overcome these physical constraints are the ability to extract accurate measurements at close proximity while still having visual access to the entire length of the bearing without shifting the UAVs position and to minimally achieve the current inspection method's accuracy requirement of ±3 mm. To modernize this process, we proposed a framework that extracts displacement measurement from a 3-D point cloud. This point cloud is generated with stereoscopic images captured from a novel cascaded bidirectional linear actuator pseudostereo camera system installed in a UAV. The framework serves to be a more accurate and safer inspection method.

Most 3-D measurement algorithms can be categorized into the following two groups: noncontact and contact. Contact methods primarily involve a coordinate-measuring machine [4], but these methods have limited measurement range and are too bulky to install on small UAVs. Meanwhile, noncontact methods, including laser-based scanning [5], [6], structured light [7], and stereo vision [8], [9], have been thoroughly explored with some reaching nano-scale. While laser-based methods are the most
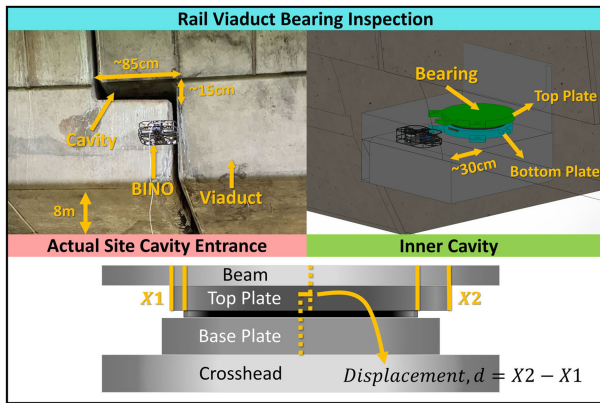
Fig. 1. BINOv2 navigating into an actual rail viaduct cavity (top left), with illustration of bearing within the cavity (top right), and bearing schematic to highlight measurement requirement (bottom).

accurate, but they typically require multiple hardware components and a complicated setup. Next, LiDAR [6] and structured light sensors, though popular payloads on UAVs, would still need the bulk of BINOv2's space. Without moving the sensors to cover the bearing's curvature, it is impossible to construct a quality geometric profile. Vision-based may be the most feasible method as it is achievable using a monocular camera, presenting the smallest form factor among the discussed methods.

Ren *et al.* [8] and Xia *et al.* [9] both use stereo vision to obtain submillimeter 3-D measurements of small mechanical parts. Ren [8] use curve-fitting techniques where linear features are extracted at different image scales to form the desired nonlinear geometrical feature amidst a noisy background. Xia [9] uses the Canny algorithm [10] to extract the hole's edge points then apply an ellipse-fit with epipolar constraints to obtain 3-D measurements through triangulation. Although both deliver impressive results, they are currently limited to detecting clear ellipse when the entire geometry is entirely within the image frame. That said, both papers inspired us to use the bearing's geometry for curve-fitting processes as it is cylindrical, but the challenge is the camera can only access the bearing's orthographic frontal view.

Conventionally bearing stereoscopic images are obtained by deploying two or more cameras around the circumference for photogrammetry processes, either for 3-D measurements or reconstruction [11]. However, the UAV's limited payload and small cavity space disallow numerous cameras with a fixed baseline onboard. Even with multiple cameras within the UAV's chassis, the combined field of view (FOV) might still be insufficient due to the usual close proximity landing position between BINOv2 and the bearing since BINOv2's width is shorter than the bearing. Another concern would be that the fixed camera's baseline distance will not obtain a stereoscopic image with sufficient overlap leading to poor reconstruction. Adopting a fish-eye lens is a possible solution to increase the individual camera's FOV [12]. However, this leads to high distortion of the images, resulting in poor reconstruction too. The minimalistic

setup would be to use a single camera to achieve pseudostereo. Some methods to achieve pseudostereo use mirrors to achieve overlapping scene to form the stereoscopic image [13]. However, the overall camera's FOV range with this setup will remain limited to a single stereoscope image from a single position.

We present an accurate framework for our core contribution to extract displacement measurement of a rail viaduct bearing using the 3-D reconstruction point cloud generated with stereoscopic images captured from a cascaded linear actuated pseudostereo camera system in a single sequence (see Fig. 2). With region projection through a curve-fitting technique to patch the bearing's point cloud's missing region, our method can obtain the complete bearing's geometric profile with the camera having only access to the orthographic front view. Furthermore, our cascaded rack-and-pinion linear actuator effectively extends the overall FOV of the camera beyond the length of BINOv2 by $(n+1)$ times the base rack's length, where n is the number of rack tiers. This design retains a small form factor allowing the entire actuator system to fit within the frame of BINOv2. The extended FOV also introduces rotation robustness to the extracted displacement accuracy regardless of BINOv2 rotated landing position w.r.t the bearing.

II. CUSTOM PSEUDOSTEREO CAMERA SYSTEM

The custom camera system design features a monocular camera installed on a cascaded bidirectional rack-and-pinion linear actuator, emulating the characteristic of a parallel stereo camera. Traditional rack-and-pinion usually consists of a single tier of one rack and one pinion and a single motor to drive it. In that setup, the pinion's maximum travel distance is limited to the rack's length. Within a confined and limited operating space such as our proposed application, there are a few situations where the camera assembly's useful FOV might be insufficient to cover the entire inspection target's length, rendering any measurement algorithm useless. These situations include the UAV landing too near or toward the cavity's sides, which were major pitfalls in our previous work [3]. Due to the tight space within the cavity, it is challenging to precisely maneuver or hold a stable hover inside. It is more practical to design the camera assembly such as its FOV is sufficient regardless of landing position.

Another benefit of an extended FOV is that the inspection target's side profile is more accessible for capture, leading to a fuller detailed point cloud model. Overall, this helps the curve-fitting accuracy at the subsequent projection step in the framework, ultimately improving the displacement extraction accuracy. Therefore, the cascaded actuator ensures that the point cloud's curvature obtainable first maximizes before the projection occurs. Fig. 3 illustrates the point cloud of the proposed inspection target generated from different configurations. As seen in the figure, rotating the camera results in an warped and unusable point cloud. The point cloud generated with the cascaded linear actuator is closer to a semicircle curvature in its top orthographic view than the standard rack-and-pinion linear actuator used in our previous work [3], where the point cloud's curvature is not as extensive.
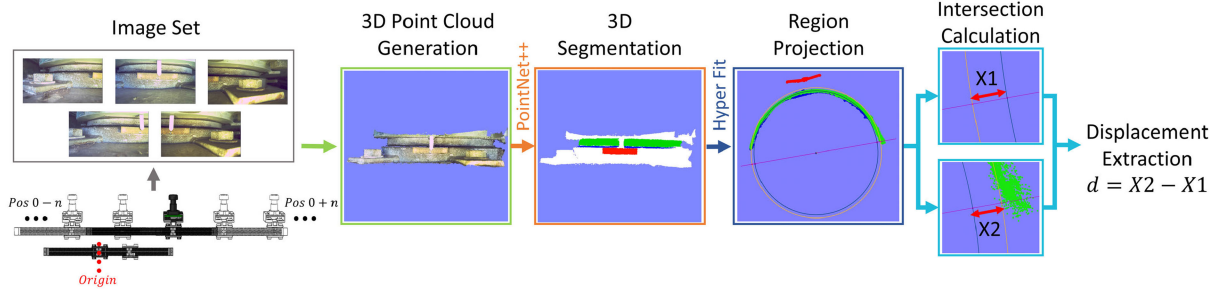
Fig. 2. Overall displacement extraction algorithm flowchart, coupled with custom cascaded bidirectional linear actuator.
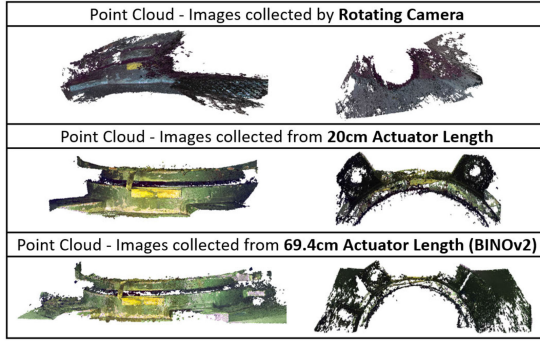

Fig. 3. Comparing completeness of point cloud generated by different configuration. Frontal view (left), top view (right).

A. Cascaded Bidirectional Linear Actuator Design

The proposed custom linear actuator features a scalable, small form factor. In this bidirectional, cascaded design, the pinion's overall travel distance is n times the rack's length, depending on the number of tiers, n, and without adding any additional motor driving it. The displacement relationships of the nth tier rack, $x_{r(n)}$, and pinion, $x_{r(n)}$, w.r.t the base pinion, $x_{p(1)}$ are $x_{p(n)} = (2n - 1)x_{p(1)}$ and $x_{r(n)} = 2(n - 1)x_{p(1)}$. Based on the relationships described, the FOV of the camera, which is mounted on the nth pinion, scales up by a multiple of n.

Cables drive the individual pinion on individual rack on both sides. Only one stepper motor drives the whole setup through the first or base rack's pinion block and, in turn, tension the cables to drive the rest pinion blocks to form the cascaded design (see Fig. 4). Reducing the number of required driving actuator to one minimized the form factor of the overall actuator design eliminating the need for any other driver boards or motors to drive the additional tiers of racks. An added benefit is that the base motor's torque τ_m required to drive the entire cascaded actuator is lower in the proposed tiered cable driven design than having an individual motor on each rack. Derivation of the static mathematical model of the cascaded actuator begins from the top nth tier pinion. Through the free body diagram [see Fig. 4(a)], the cable tension force acting upon the nth pinion is

$$F_{p(n)} - f_{r(n)} = m_{p(n)}\ddot{x}_{r(n)} = 0 \tag{1}$$

$$T_{p(n)} = (\mu_s g m_{p(n)})/\cos\theta \tag{2}$$

where $F_{p(n)}$ is the force acting upon the nth pinion, $f_{r(n)}$ is the friction force between the nth rack and pinion, which is also equal to $\mu_s m_{p(n)}g$ whereby μ_s is the static friction coefficient, $m_{p(n)}$ is the nth pinion mass. $T_{p(n)}$ represents the tension of cable driving the pinion, θ is the cable angle at the anchor points on the pinion, and $\ddot{x}_{r(n)}$ is the acceleration of the nth rack. Subsequently, using the derivation of the nth pinion model, and through the free body diagram [see Fig. 4(b)], derivation of the nth rack static model is

$$F_{r(n)} + f_{p(n)} - f_{p(n-1)} - 2F_{p(n)} = m_{r(n)}\ddot{x}_{p(n-1)} = 0 \tag{3}$$

$$F_{r(n)} = \mu_s g(m_{r(n)} + 2m_{p(n)}) \tag{4}$$

where $F_{r(n)}$ is the force acting upon the n_{th} rack, $m_{r(n)}$ is the nth rack mass, and $\ddot{x}_{r(n)}$ is the acceleration of the $(n-1)_{th}$ rack. Carrying on with the next $(n-1)_{th}$ pinion, the static derivation [see Fig. 4(c)] of the forces is

$$F_{p(n-1)} + F_{p(n)} + f_{r(n)} - F_{r(n)} - f_{r(n-1)} = 0 \tag{5}$$

$$T_{p(n-1)} = \left[\mu_s g(m_{p(n)} + m_{r(n)} + m_{p(n-1)})\right]/\cos\theta. \tag{6}$$

A derivation pattern of the tension force acting upon the pinion n_{th} and $(n-1)_{th}$ is identified whereby the resolved forces are a function of the friction force of the pinion and a summation of the masses above itself. Therefore, the overall driving force acting upon the base pinion can be generalize to

$$F_{\text{drive}} = \frac{\mu_s g}{\cos\theta}\left[\sum_{i=2}^{n}(m_{p(i)} + m_{r(i)}) + m_{p(1)}\right] \tag{7}$$

where $m_{p(i)}$ is the mass of the pinion block from the second tier to nth tier, the rack masses $m_{r(i)}$ from the second rack to nth tier, $m_{p(1)}$ is the mass of the first pinion block

$$F_{\text{drive}} = \frac{\tau_p}{r_p} = \frac{\eta_{mp}\tau_m}{r_{mp}r_m}. \tag{8}$$

Transmission torque from the motor gear, τ_m, to the base pinion gear, τ_p, can be modeled as in (8) where η_{mp} represents power transmission efficiency, and r_{mp} is the ratio of motor gear radius r_m and pinion gear radius r_p. Therefore, the motor torque at the motor gear needs to be greater the minimum force requirement of the custom cascaded actuator as shown in

$$\tau_m \geq \frac{r_{mp}r_m\mu_s g\left[\sum_{i=2}^{n}(m_{p(i)} + m_{r(i)}) + m_{p(1)}\right]}{\eta_{mp}\cos\theta}. \tag{9}$$
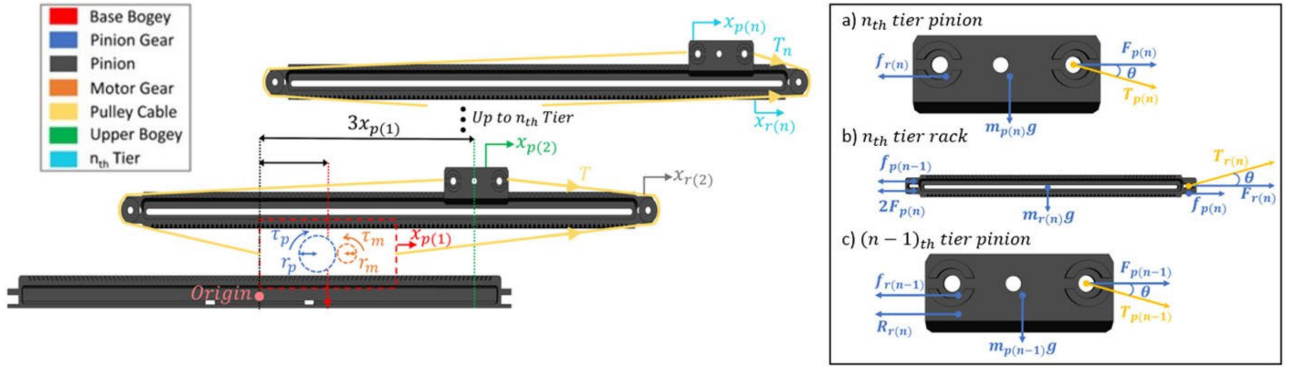
Fig. 4.   Free-body diagram of the custom bidirectional cascaded linear actuator.
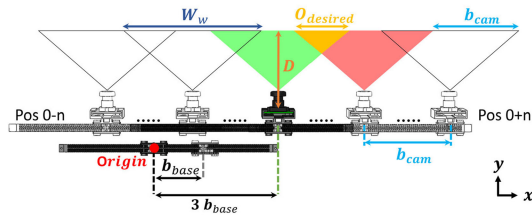

Fig. 5.   Overview of pseudostereo camera configuration featuring the custom cascaded bidirectional actuator.

The cascaded linear actuator used in our proposed application consists of two tiers. The entire system is 3-D printed using nylon material. From the material datasheet, the material onyx's static friction coefficient, μ_s, is approximately 0.2. The motor gear radius r_m is $0.0052\,m$ with a ratio, r_{mp}, of $1.625:1$ to the pinion gear, with a conservation assumption of 0.6 for transmission efficiency, η_{mp}. Lastly, the summation of the racks and pinions masses $m_{p(n)}$ and $m_{r(n)}$ is $0.167\,\mathrm{kg}$, while the cable angle, θ, is $30°$ at its max. Hence, using (9), the minimum motor torque τ_m multiple with a torque safety factor of 2.25 is $0.01028\,\mathrm{N\cdot m}$, which is the minimum requirement to drive the cascaded linear actuator.

B. Pseudostereo Camera Configuration

A parallel stereo camera has two cameras along a parallel optical axis, and the captured image pair would only have a translational difference. Our proposed set-up emulates the parallel stereo camera by capturing image pairs from various positions on the actuator based on a baseline distance b_{base}.

Broadly, the overall camera assembly operation and our parallel pseudostereo camera setup calibration process remain the same as our previous work [3]. However, the translation from a standard single-tiered rack-and-pinion to an nth tiered cascaded set-up changes the baseline calculation. In order to achieve desired overlap percentage O_{desired}, between images for photogrammetry [14], the baseline distance b_{base}, which represents the distance where the base pinion needs to travel for each image

(see Fig. 5), is determined by

$$b_{\mathrm{base}} = [W_w \times (1 - O_{\mathrm{desired}})]/(2n - 1) \qquad (10)$$

where W_w represents the horizontal camera FOV. The maximum horizontal FOV, W_{max}, of the camera in this pseudostereo configuration is given as $W_{\mathrm{max}} = L_{\mathrm{max}} + W_w$, where L_{max}, which is the maximum effective travel length of the nth tier pinion of the cascaded linear actuator. Therefore, the L_{max} for our proposed application, which consists of a two tiered configuration, is twice the length of the base rack L_{base}.

The mechanical backslash that the cascaded linear actuator may face is taken into account. Even with the maximum backslash error of ± 5 mm and at the minimum distance between the camera and bearing of 70 mm, the image overlap percentage is ensured to be still be minimally 70%.

III. DIMENSION EXTRACTION ALGORITHM PIPELINE

The proposed dimensional extraction algorithm consists of three main elements within its pipeline: Point cloud dataset preparation, 3-D semantic segmentation, and Dimensional displacement extraction through projection. When developing the algorithm, the primary considerations were producing a scalable and accurate end to end solution for straightforward implementation across the thousands of inspection sites.

A. Data Generation and Population Pipeline

Training a highly accurate neural network model is dependent on the quality of the training dataset in terms of its variance, quantity, and accurate representation of the target of interest. Obtaining such a dataset is administratively expensive and challenging due to the target's remote location. The wide range of different bearing series would mean a large dataset is needed to cover all cases. Hence, instead of collecting on-site images to generate enough point clouds to form the whole training dataset, our proposed pipeline only needs a small sample of point clouds generated from image sets. After that, various affine and colorspace augmentation methods populate the rest of the training dataset.

The dataset generation pipeline involves first obtaining a point cloud model from an image set of the actual object, which is the
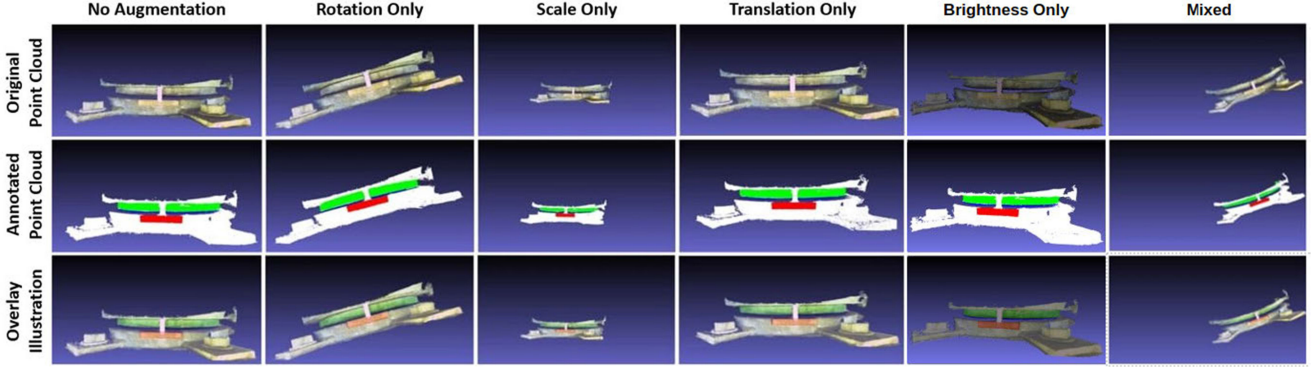
Fig. 6. Training dataset consists of the original bearing point cloud (left) and different point cloud affine and colorspace augmentation, chosen at random that includes, scale, rotation, translation, brightness, and a mixture.

bearing in our proposed application. Before the images captured are utilized to generate the point cloud, they are undistorted using the derived camera parameters from the OpenCV [15] monocular camera calibration process. The generation of point cloud is done using the python library of the Metashape software by Agisoft LLC.[1] The proprietary process of Agisoft uses SIFT algorithm [16] as an initial step to extract important feature points of the stereoscopic image pair. The scale-invariant nature of the algorithm helps resolve issues such as scale ambiguity. Next, the generated point cloud is annotated using an open-source annotation tool semantic segmentation editor.[2] After generating all available real point cloud data, the rest of the point clouds that form the bulk of the training dataset are populated through an augmentation process using a custom script. The first step of augmentation is to translate the point cloud generated from real data, $P = \{p_x, p_y, p_z \in \mathbb{R}\}$, to a centralized position P_{ctr} by subtracting every point coordinate with its individual axes' mean value

$$P_{\text{ctr}} = \{p_x, p_y, p_z \in \mathbb{R} | (\forall p_x - \overline{p_x}), (\forall p_y - \overline{p_y}), (\forall p_z - \overline{p_z})\}. \tag{11}$$

Subsequently, we normalize the centralised point cloud into a unit circle P_{norm}, where

$$P_{\text{norm}} = \{p_x, p_y, p_z \in \mathbb{R} | \frac{\forall (p_x, p_y, p_z)}{\max_x(\sqrt{[b]\sum_i |x|_i^2})}\}. \tag{12}$$

Finally, randomly select an augmentation method using a uniform distribution. The augmentation methods consist of six different types and it is described as

$$\zeta_i \in [\zeta_i^{\text{lb}}, \zeta_i^{\text{ub}}], i = \text{Prob}(1|2|...N) \tag{13}$$

where ζ is the augmentation method, and the strength of the augmentation is within and lower bound, lb and upper bound, ub, and N is the number of augmentation method for selection. The methods include no augmentation applied, rotation only, translation only, scale only, brightness only, and mixed (see Fig. 6). The selection of which axes of the point cloud to apply the augmentation and augmentation strength is also uniformly

[1][Online]. Available: https://www.agisoft.com/
[2][Online]. Available: https://github.com/Hitachi-Automotive-And-Industry-Lab/semantic-segmentation-editor/

random selected. The training dataset consists of 550 point clouds, where 530 is for the training (30 real point cloud with 500 synthetic point cloud) and 20 for the validation stage (all real point cloud).

B. Neural Network Architecture and Training Results

Point clouds have an irregular format that possesses unorderedness, meaningful subsets formed by points' interaction, and various transform invariance. Most processes involved transforming the point cloud data to regular voluminous grids or collections of images before feeding them into typical convolutional architectures. However, this often produces resulting data being voluminous and introduces quantization effects that may obscure the data's natural invariances. Therefore, to retain an accurate Euclidean representation of the point cloud, the network architecture needs to ingest point cloud models as input without modifications directly. While the model's architectural design has not been modified, the hyperparameters have been extensively tuned to obtain the best metrics. Furthermore, the training pipeline's data ingestion portion has been modified to take in point cloud PCD format directly.

The PointNet++[17], a hierarchy network architecture whose concept runs PointNet [18] recursively and progressively through an increasing scale. Its key differentiator from other 3-D semantic segmentation network architecture is that it consumes labeled point cloud data directly as input, which helps us to overcome irregular point cloud format issues. The network architecture employed consists of a four-hierarchical level, with each level featuring a sampling layer, grouping layer, and pointnet layer. The number of points chosen for the sampling layer is 32 across all levels, the number of points for K nearest neighbor in the grouping layer is [1024, 256, 64, 16], and the clustering radius for the pointnet layer is [0.05, 0.1, 0.2, 0.4]. The optimizer is Adam, with an learning rate, $\alpha = 0.0015$. All models are trained on an Nvidia RTX 2080 Ti GPU card.

The metrics for evaluating the model's performance are the total loss, accuracy score, mean intersection of union score (mIOU Score), recall, precision, and $F1$ score. The resultant metrics of the training and validation processes are shown in Table I.

TABLE I
MODEL TRAINING RESULTS

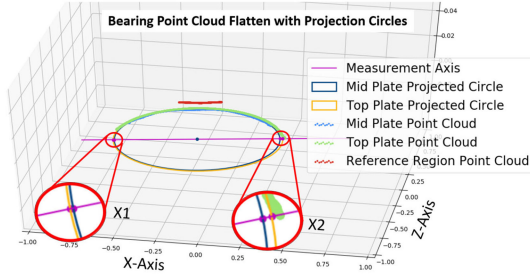|	Loss	mIOU	Acc	Recall	Precision	F1 Score
Train	0.0499	0.9106	0.9362	-	-	-
Validate	0.0994	0.8520	0.8688	0.8622	0.8711	0.8666

Fig. 7. Illustration of the individual point clouds, projection of best-fit circles, measurement axis, and displacement extraction.

C. Point Cloud Projection and Measurements Extraction

The critical information needed for extraction for the rail viaduct bearing inspection process is the displacement between the top and the base plate [see Fig. 1(bottom)]. However, the point cloud region at the semicircle area, where dimensions need to be measured to calculate the displacement, is often noisy or missing. The linear actuator's images are ineffective in capturing the bearing's details beyond 50% of the circumference. Hence, before the displacement information is extractable, a projection of the rest of the bearing's circumference is needed. By reducing the 3-D problem into a 2-D space, using only the x and z axes points of the point cloud $P_{xy} = \{p_x, p_y \in \mathbb{R}\}$, the bearing's full circumference projection is made through finding the best-fit circle. Both the bearing's top and the base plate, isolated by neural network model segmentation, each has its circle projection. The Euclidean distance between both circles at each side of the semicircle region forms is the displacement calculation, which is $d = X1 - X2$, where $X1$ and $X2$ are the Euclidean distance of the left and right side, respectively (see Fig. 7).

The proposed algorithm first filters the neural network segmentation model's prediction to extract the individual point clouds of region of interest, namely the top, the base plate of the bearing, and a reference region. After that, a clustering algorithm, dbscan, [19] removes any outliers in the individual base and top plate point cloud. Next, our proposed algorithm uses a technique known as hyper least square or hyperfit [20] to calculate the best-fit circles for the segmented individual point cloud. There are two methods type for circular fitting that include geometric and algebraic. Based on an analysis done on circular fitting algorithms [21], the hyper fit technique, which is translation and rotation invariant, yields the best result by combining two other methods Pratt [22] and Taubin [23], to eliminate essential bias. This result in best-fitted circles, $O_{(b|t)}(\text{ctr}_x, \text{ctr}_z, \text{rad})$, where ctr_x and ctr_z are the x, z coordinate of the fitted circles' center and rad is its radius. Then, the length of the consolidated reference region points ref_l, in the scale of

a unit circle, is calculated by finding the distance between the points of edges using

$$\text{ref}_l = \sqrt{(dx_{\text{ref}})^2 + (dz_{\text{ref}})^2} \quad (14)$$

where dx_{ref} and dz_{ref}, which is equal to $\max(\text{ref}_{(x,z)}) - \min(\text{ref}_{(x,z)})$, are the distance between the maximum and minimum of x, z axes coordinates in the reference region individual point cloud.

To obtain an accurate measurement of the best-fit circles' displacement at the semicircle area, a line projection parallel to the reference region and that cuts through the circles' center is used as a measurement axis. The intersection points of the circles and the line then serve as coordinates to calculate the displacement. The gradient of this line projection, m, uses a least-square first-degree polynomial fit of the reference region points which follows the form of $Y = a + bX$. Subsequently, the intersection points between best-fit circles and the line projection are determined using the quadratic formula

$$a = dx^2 + dz^2 = (p_{rx} - p_{lx})^2 + (p_{rz} - p_{lz})^2 \quad (15)$$
$$b = 2[dx(p_{lx} - \text{ctr}_{(b|t)x}) + dy(p_{lz} - \text{ctr}_{(b|t)z})] \quad (16)$$
$$c = (p_{lx} - \text{ctr}_{(b|t)x})^2 + (p_{lz} - \text{ctr}_{(b|t)z})^2 - \text{rad}_{(b|t)}^2 \quad (17)$$

where a, b, and c are the variables. The left and right edge coordinates of the line projection, $p_{l(x,z)}$ and $p_{r(x,z)}$ are equal to $(-1, [m(-1 - \text{ctr}_{bx})] + \text{ctr}_{bz})$ and $(1, [m(1 - \text{ctr}_{bx})] + \text{ctr}_{bz})$, respectively. The intersection points of the best-fit circles' left and right sides for the base and top plate and the line projection, $\text{lhs}_{(b|t)}$ and $\text{rhs}_{(b|t)}$, is

$$(\text{lhs}_{(b|t)x}, \text{lhs}_{(b|t)z}) = (dx \times \text{rt}_1 + p_{lx}, dz \times \text{rt}_1 + p_{lz}) \quad (18)$$
$$(\text{rhs}_{(b|t)x}, \text{rhs}_{(b|t)z}) = (dx \times \text{rt}_2 + p_{lx}, dz \times \text{rt}_2 + p_{lz}) \quad (19)$$

where rt_1 and rt_2 are the roots of the quadratic equation. Lastly, the calculation of the Euclidean distance between both the best-fit circle is then

$$X1 = \left[(\sqrt{(\text{lhs}_t - \text{lhs}_b)_x^2 + (\text{lhs}_t - \text{lhs}_b)_z^2}) / \text{ref}_l \right] (\text{ref}_a) \quad (20)$$

$$X2 = \left[(\sqrt{(\text{rhs}_t - \text{rhs}_b)_x^2 + (\text{rhs}_t - \text{rhs}_b)_z^2}) / \text{ref}_l \right] (\text{ref}_a) \quad (21)$$

where ref_a is the actual dimensional length of the reference region. The need for the actual reference length scales the overall point cloud from a unit circle to the real world coordinates whereby an accurate measurement of displacement is obtainable using $X1 - X2$. Any object that has a dimension that does not change over time will make an ideal reference region. Some reference regions may include the bearing's base or the built-in scale bar used for this article. The actual dimension of the reference region is available from the bearing schematic. Most bearing scale bars reading fades off over time, making it unusable for the current inspection process, but our algorithm does not depend on the reading.
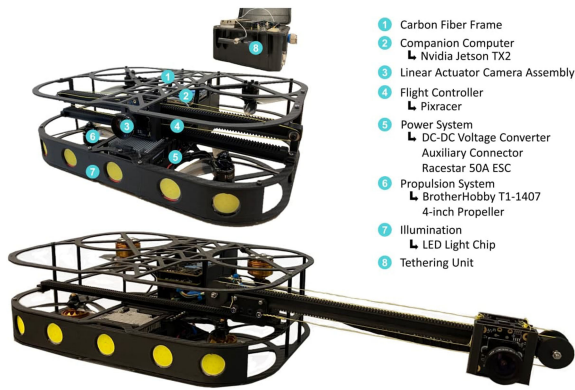
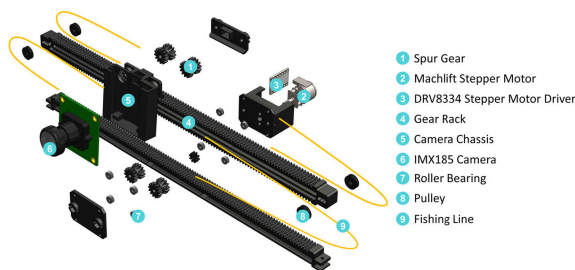Fig. 8. Components breakdown of the custom BINOv2 UAV.

Fig. 9. Components breakdown of cascaded bidirectional linear actuator camera assembly design.
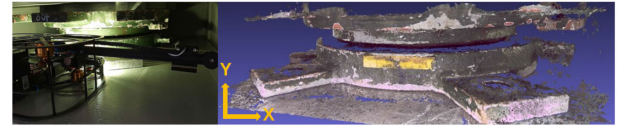

Fig. 10. BINOv2 within mock-up cavity to capture stereoscopic images of decommissioned bearing.

TABLE II
FULL OPERATION TRIAL EXPERIMENTAL RESULTS—DESCRIPTION AND ERRORS

Displacement Measurement (mm)					
Ground Truth	6.000	**Mean**	6.354	**Max Error**	0.753
Standard Deviation	0.546	**RMSE**	0.504		

IV. EXPERIMENTS AND RESULTS

A. Unmanned Aerial Platform Description

The dimension of the custom BINOv2 UAV platform currently used for the proposed application is $265 \times 76 \times 200$ mm. The camera assembly's maximum length L_{max}, which features a two-tier cascaded linear actuator, is 510 mm. BINOv2 is slightly wider than its predecessor, where it was 231 mm. However, BINOv1 only had a maximum actuator length of 177.53 mm compared to the current version, 510 mm length, while only sacrificing 30 mm in its overall length. The viaduct bearing entrance's overall size is $850 \times 150 \times 300$ mm, and this dimension is sufficient for BINOv2 to enter the cavity with ease. The BINOv2's conceptual design retains the box-in design where it ensures that the components within the frame are well protected even if the UAV brush against any side of the wall. The camera used in the camera assembly is the IMX185 with an M12 lens. L_{max} with the camera renders an entire horizontal camera FOV, W_{max}, of 694 mm. Figs. 8 and 9 show the component breakdown of BINOv2 and the cascaded linear camera assembly, respectively.

B. Mock-Up Viaduct Experiment

The remote and inaccessible location of the bearing prevents various controlled experiments that require human access. It is, for example, not logistically possible to compare our proposed algorithm with other state-of-the-art commercial stereo cameras or other methods as the only medium to access the bearing is

through the described UAV platform. Therefore, to overcome this obstacle and still conduct realistic and accurate experiments, we designed and built a dimensionally exact mock-up cavity (see Fig. 10) to perform different experiments to validate the performance and viability of our proposed framework. There is a real decommissioned rail viaduct bearing within the mock-up cavity, precisely the same as those installed at the actual site, for realism. The whole mock-up cavity is installed on a scissor lift with a peak height of 2 m. Ground truth measurement of displacement is first obtained by lowering the lift. For all experiments, the mean and standard deviation of the displacement measurements provide insight into the algorithm's consistency. The root mean squared error (RMSE) and max error demonstrate the algorithm's accuracy. There are four separate experiments conducted on the mock-up cavity.

1) Full Conduct Inspection Operation: The first part involves conducting the full inspection operation to simulates the inspection conduct at actual on-site operation. A UAV pilot manually operates BINOv2 to take off from the ground, enters the mock-up cavity, jacked to its highest height of 2 m, and lands uncontrolled within the cavity, and executes the algorithm to extract the bearing's dimensions. We carried out a total of ten repeated flights for this experiment. Table II provides the consolidation of the experiment result from these ten flights. The average RMSE of the ten runs is submillimeter at 0.504 mm, which fulfills the current bearing inspection methods' accuracy requirement of ± 3 mm.

2) Robustness Against Rotated Landing Positions: The second part involves lowering the lift and placing BINOv2 at various rotated landing positions w.r.t the cavity's bearing to test the proposed algorithm's robustness rotated landing position. This robustness is critical as the landing position of the UAV is highly uncontrolled during normal operation, and the algorithm needs to produce accurate measurements regardless of the UAV's position. The second experiment consists of ten execution of the algorithm for each of the rotated positions. Table III provides the consolidation of the experiment result for the individual position. The max average RMSE of the ten execution along all positions is 0.822 mm, which is still much within the current standards of ± 3 mm.

TABLE III
ROTATED LANDING POSITION TRIAL EXPERIMENTAL
RESULTS—DESCRIPTION AND ERRORS

Displacement Measurement (mm)						
	10°	20°	30°	330°	340°	350°
Ground Truth	6.000	6.000	6.000	6.000	6.000	6.000
Standard Deviation	0.792	0.892	0.723	0.465	0.320	0.658
Mean	5.812	6.323	6.765	6.589	6.902	6.109
RMSE	0.797	0.757	0.635	0.801	0.822	0.539
Max Error	1.233	0.989	1.168	1.212	1.235	1.104

TABLE IV
CASCADED VERSUS TRADITIONAL ACTUATOR VERSUS PREVIOUS WORK
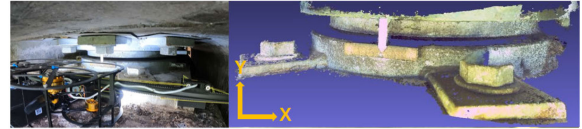EXPERIMENTAL RESULTS—DESCRIPTION AND ERRORS

Displacement Measurement (mm)			
Algorithm	Proposed Framework		Previous Work [3]
Actuator Type	Cascaded	Traditional	Traditional
Ground Truth	6.000	6.000	6.000
Standard Deviation	0.256	0.398	1.175
Mean	6.222	4.028	5.727
RMSE	0.485	1.933	1.206
Max Error	0.917	2.761	2.755

TABLE V
PROPOSED ALGORITHM VERSUS MEASUREMENTS FROM COMMERCIAL
STEREO CAMERAS—DESCRIPTION AND ERRORS

Displacement Measurement (mm)			
Camera Type	Monocular	Intel RealSense 435i	Structure Core
Minimum Distance	70mm	250mm	390mm
Ground Truth	6.000	6.000	6.000
Standard Deviation	0.558	1.011	1.043
Mean	6.380	1.491	-0.155
RMSE	0.601	4.509	6.155
Max Error	1.012	5.786	8.022

3) Cascaded Versus Traditional Linear Actuator's Impact on Measurement Accuracy: The third part compares displacement measurement obtained from the cascaded linear actuator, traditional single-tiered linear actuator, and the 2-D measurement algorithm from our previous work [3] when the camera is directly parallel to the orthographic frontal view. The third experiment consists of ten execution of the algorithm, each for the cascaded and traditional linear actuator and the 2-D measurement algorithm from our previous work. Table IV provides the consolidation of the experiment results for the third experiment. The results show that measurements extracted using the cascaded linear actuator with the proposed algorithm are significantly more accurate in all metrics than those obtained through a traditional linear actuator and algorithm from our previous work.

4) Proposed Algorithm Versus State-of-the-Art Stereo Cameras: The fourth part compares the measurement's quality extracted from the proposed algorithm and two state-of-the-art popular commercial stereo cameras at the minimum distance to the bearing. Similarly to the previous experiments, ten execution of each method is carried out. Measurements extracted from the point cloud generated by RealSense and structure core are done manually, selecting the points to ensure the most accurate measurements. Table V provides the consolidation of the experiment results for the third experiment. For the structure core, the minimum distance to obtain a usable point cloud is beyond the cavity space's available length. For both the stereo cameras,

Fig. 11. BINOv2 within the actual site cavity to capture stereoscopic images of bearing.

TABLE VI
ACTUAL SITE TRIAL EXPERIMENTAL RESULTS—DESCRIPTION AND ERRORS

First Site Experiment Displacement Measurement (mm)					
Ground Truth	4.000	Mean	4.290	Max Error	1.520
Standard Deviation	0.867	RMSE	0.925		
Second Site Experiment Displacement Measurement (mm)					
Ground Truth	-0.500	Mean	0.264	Max Error	1.554
Standard Deviation	0.602	RMSE	0.974		
Third Site Experiment Displacement Measurement (mm)					
Ground Truth	2.000	Mean	2.180	Max Error	1.498
Standard Deviation	0.840	RMSE	0.644		

the measurements extracted significantly underperformed as compared to our proposed algorithm. Neither stereo cameras can fulfill the minimum accuracy requirement of ±3 mm.

C. Actual Viaduct Site Trial

For the actual site trial, we selected three different sites with the same series of bearing. All three sites' cavity was roughly 8 m high, and we conducted five full flight inspection operations for each site (see Fig. 11). There are two criteria for selecting the trial site. The first prioritizes the bearings' condition, where the readings on the scale bar preinstalled are still clear and visible. This requirement is to facilitate the evaluation of the algorithm's performance by comparing it with the ground truth measurement on the scale bar. The second requires the site bearing to be the same series as the decommissioned bearing in the mock-up cavity as all share the same trained 3-D segmentation model. The algorithm executes twice each flight, producing a collection total of ten displacement measurement results for each site. While the pilot did try to land BINOv2 as parallel to the viaduct bearing as possible, the landing position within the cavity for each flight is uncontrolled and is different for every flight. Table VI provides the consolidation of the experiment result for all the experimental trial site. Results show that the RMSE values from all three sites are below 1 mm, which is better than the current accuracy standard of ±3 mm.

V. CONCLUSION

Our proposed framework has demonstrated the capability of using our algorithm to extract accurate displacement measurement from a rail viaduct bearing using an image set collected from a cascaded linear camera assembly in a single sequence, for both experimental mock-up and actual operation site environment. Furthermore, our algorithm is robust against landing positions inconsistencies within a cavity, shown in the consistency of the RMSE values in all experiments. To further improve displacement measurement accuracy, the optimal step would be to eliminate the need for best-fit circles projection to obtain the

complete geometric profile of the bearing, and this would require the curvature of the point cloud to cover at least or more than 50% of the circumference. Our future work will address the saturation problem where the bearing's side profile beyond the semicircle region cannot be seen anymore due to the actuator's linear motion constraint.

REFERENCES

[1] L. Wang, Z. Zhang, and X. Luo, "A two-stage data-driven approach for image-based wind turbine blade crack inspections," *IEEE/ASME Trans. Mechatronics*, vol. 24, no. 3, pp. 1271–1281, Jun. 2019.

[2] H. Yu, W. Yang, H. Zhang, and W. He, "A UAV-based crack inspection system for concrete bridge monitoring," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2017, pp. 3305–3308.

[3] D. Lee, J. Liu, S. M. Lee, and S. Foong, "Automated dimensional extraction of different regions using single monocular camera in pseudo-stereo configuration," in *Proc. IEEE/ASME Int. Conf. Adv. Intell. Mechatronics*, 2020, pp. 314–321.

[4] X. Wang, J. Xian, Y. Yang, Y. Zhang, X. Fu, and M. Kang, "Use of coordinate measuring machine to measure circular aperture complex optical surface," *Measurement*, vol. 100, pp. 1–6, 2017.

[5] L. Clark, B. Shirinzadeh, Y. Tian, and D. Oetomo, "Laser-based sensing, measurement, and misalignment control of coupled linear and angular motion for ultrahigh precision movement," *IEEE/ASME Trans. Mechatronics*, vol. 20, no. 1, pp. 84–92, Feb. 2015.

[6] H. C. Jo, J. Kim, K. Lee, H.-G. Sohn, and Y. M. Lim, "Non-contact strain measurement for laterally loaded steel plate using lidar point cloud displacement data," *Sensors Actuators A, Phys.*, vol. 283, pp. 362–374, 2018.

[7] A. N. André, P. Sandoz, B. Mauzé, M. Jacquot, and G. J. Laurent, "Sensing one nanometer over ten centimeters: A microencoded target for visual in-plane position measurement," *IEEE/ASME Trans. Mechatronics*, vol. 25, no. 3, pp. 1193–1201, Jun. 2020.

[8] Z. Ren, J. Liao, and L. Cai, "Three-dimensional measurement of small mechanical parts under a complicated background based on stereo vision," *Appl. Opt.*, vol. 49, pp. 1789–1801, Apr. 2010.

[9] R. Xia *et al.*, "An accurate and robust method for the measurement of circular holes based on binocular vision," *Meas. Sci. Technol.*, vol. 31, Nov. 2019, Art. no. 025006.

[10] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[11] S. Maehara, K. Nitta, and O. Matoba, "Three-dimensional measurement and imaging based on multicameras randomly distributed on the circumference," *Appl. Opt.*, vol. 47, pp. 594–601, Feb. 2008.

[12] H. Kim, J. Jung, and J. Paik, "Fisheye lens camera based surveillance system for wide field of view monitoring," *Optik*, vol. 127, no. 14, pp. 5636–5646, 2016.

[13] T. P. Pachidis and J. N. Lygouras, "Pseudo-stereo vision system: A detailed study," *J. Intell. Robot. Syst.*, vol. 42, no. 2, pp. 135–167, 2005.

[14] D. Wierzbicki, "Multi-camera imaging system for UAV photogrammetry," *Sensors*, vol. 18, Jul. 2018, Art. no. 2433.

[15] G. Bradski, "The OpenCV library," *Dr Dobb's J. Softw. Tools*, vol. 120, pp. 122–125, 2000.

[16] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, 1999, vol. 2, pp. 1150–1157.

[17] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 5105–5114.

[18] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation," *IEEE Conf. Comput. Vision Pattern Recognit. (CVPR)*, vol. abs/1612.00593, pp. 77–85, 2017, doi 10.1109/CVPR.2017.16.

[19] E. Martin *et al.*, "A density based algorithm for discovering density varied clusters in large spatial databases with noise," *Int. J. Comput. Appl.*, Kdd, vol. 96, no. 34, pp. 226–231, 1996.

[20] K. Kanatani and P. Rangarajan, "Hyper least squares fitting of circles and ellipses," *Comput. Statist. Data Anal.*, vol. 55, no. 6, pp. 2197–2208, 2011.

[21] A. Al-Sharadqah and N. Chernov, "Error analysis for circle fitting algorithms," *Electron. J. Statist.*, vol. 3, pp. 886–911, 2009.

[22] V. Pratt, "Direct least-squares fitting of algebraic surfaces," *Proc. Annu. Conf. Computer Graphics Interactive Tech.*, 1987, pp. 145–152.

[23] G. Taubin, "Estimation of planar curves, surfaces, and nonplanar space curves defined by implicit equations with applications to edge and range image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 11, pp. 1115–1138, Nov. 1991.
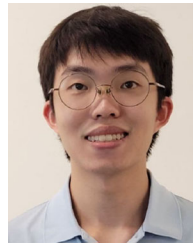
Denzel Lee received the B.Eng. (Hons) degree in aerospace systems from the Singapore Institute of Technology, Singapore, in 2017. He is currently working toward the Ph.D. degree in Engineering Product Development (EPD) with the Singapore University of Technology and Design (SUTD) Engineering Product Development (EPD) Pillar, Singapore.

His research interests include autonomous trajectory planning for unmanned aerial systems (UAS), computer vision applications for UAS, and deep learning applications primarily in reinforcement learning and computer vision.
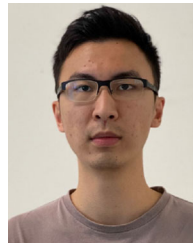
Jingmin Liu received the B.Eng. degree in aerospace systems from the Singapore Institute of Technology, Singapore, in 2019.

He is currently a Researcher Officer with the Temasek Labs (TLabs), Singapore University of Technology and Design, Singapore. His research interests include custom mechanical design of unmanned aerial systems (UAS) and integrating of external mechanical components with UAS.

Ryan Lim received the B.Eng. (Hons.) degree in engineering from the Singapore University of Technology and Design (SUTD), Singapore, in 2020. He is currently working toward the Ph.D. degree in Engineering Product Development (EPD) with the SUTD Engineering Product Development (EPD) Pillar, Singapore.

His final year thesis was on the conceptualization and development of a ground vehicle for inspecting floors of storage tanks autonomously. His research interests include reconfigurable structures, high-speed quadrotor trajectory planning, and sparse sensing for airborne applications.

Jie Lin Chan is currently working toward the B.Eng. (Hons.) degree in Information Systems Technology and Design (ISTD) with the Singapore University of Technology and Design (SUTD) Information Systems Technology and Design (ISTD) Pillar, Singapore.

His research interests include machine learning for robotics and software development in medical field such as computer vision for cells identification.

Shaohui Foong received the B.S., M.S., and Ph.D. degrees in mechanical engineering from the George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA, USA, in 2005, 2008 and 2010.

He is currently an Associate Professor with the Engineering Product Development (EPD) Pillar, Singapore University of Technology and Design (SUTD), Singapore, and a Senior Visiting Academician with the Changi General Hospital, Singapore. In 2011, he was a Visiting Assistant Professor with the Massachusetts Institute of Technology, Cambridge, MA, USA. His research interests include system dynamics control, nature-inspired robotics, magnetic localization, medical devices, and design education and pedagogy.