Geometrical-Based Displacement Measurement With Pseudo-Stereo Monocular Camera On Bi-Directional Cascaded Linear Actuator

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

Abstract—This paper 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 BINOv2: Bearing Inspector for Narrow-space Observation Version 2. BINOv2 is a tethered custom Unmanned Aerial Vehicle (UAV) 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 bi-directional 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 3D segmentation to extract targeted regions of interest. Our training pipeline generates and forms the full model's training data set 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.

Actuators, Unmanned aerial vehicles (UAVs), Condition monitoring, Displacement measurement, Learning (Artificial intelligence), Three-dimensional segmentation

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]. This paper is an extension of our previous work [2], 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 5m to 10m. However, most bearing scale bar’s 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 850mm by 150mm by 300mm. 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 UAV’s position and to minimally achieve the current inspection method’s accuracy requirement of ±3mm. To modernize this process, we proposed a framework that extracts displacement measurement from a 3D point cloud. This point cloud is generated with stereoscopic images captured from a novel cascaded bi-directional linear actuator pseudo-stereo camera system installed in a UAV. The framework serves to be a more accurate and safer inspection method.

Most 3D measurement algorithms can be categorized into two groups: non-contact and contact. Contact methods primarily involve a coordinate-measuring machine [3], but these methods have limited measurement range and are too bulky to install on small UAVs. Meanwhile, non-contact methods, including laser-based scanning [4], structured light [5], and stereo vision [6], [7], have been thoroughly explored with some reaching nano-scale. While laser-based methods are the most accurate, but they typically require multiple hardware components and a complicated setup. Next, LiDAR [4] 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

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).
Ren et al. [6] and Xia et al. [7] both use stereo vision to obtain sub-millimeter 3D measurements of small mechanical parts. Ren [6] 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 [7] uses the Canny algorithm [8] to extract the hole’s edge points then apply an ellipse-fit with epipolar constraints to obtain 3D 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 3D measurements or reconstruction [9]. 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 [10]. 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 pseudo-stereo. Some methods to achieve pseudo-stereo use mirrors to achieve overlapping scene to form the stereoscopic image [11]. 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 3D reconstruction point cloud generated with stereoscopic images captured from a cascaded linear actuated pseudo-stereo 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 Pseudo-Stereo Camera System**

The custom camera system design features a monocular camera installed on a cascaded bi-directional rack-and-pinion linear actuator, emulating 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 towards the cavity’s sides, which were major pitfalls in our previous work [2]. Due to the tight space within the cavity, it is challenging to precisely maneuver or hold stable hover. It is more practical to design the camera assembly such as its FOV is sufficient regardless of landing position.

**Fig. 3: Comparing completeness of point cloud generated by different configuration. Frontal View (left), Top View (right)**

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. Figure 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 semi-circle curvature in its top orthographic view than the standard rack-and-pinion linear actuator used in our previous work [2], where the point cloud’s curvature is not as extensive.

A. Cascaded Bi-Directional Linear Actuator Design

The proposed custom linear actuator features a scalable, small form factor. In this bi-directional, 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 \( n \)th tier rack, \( x_{r(n)} \), and pinion, \( x_{p(n)} \), 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 \( n \)th tier 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 \( \tau_m \) 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 \( n_{th} \) tier pinion. Through the free body diagram (see Fig. 4a), the cable tension force acting upon the \( n_{th} \) pinion is

\[
F_{p(n)} - f_{r(n)} = m_{p(n)}\ddot{x}_{r(n)} = 0
\]

\[
T_{p(n)} = (\mu_s g m_{p(n)})/\cos \theta
\]

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

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

\[
F_{r(n)} = \mu_s g (m_{r(n)} + 2m_{p(n)})
\]

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

\[
F_{p(n-1)} + F_{p(n)} + F_{r(n)} - F_{r(n-1)} = 0
\]

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

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 is generalize to

\[
F_{drive} = \mu_s g \sum_{i=2}^{n} (m_{p(i)} + m_{r(i)}) + m_{p(1)}
\]

\[
F_{drive} = \frac{\tau_m}{r_p} = \frac{\eta_m \tau_m}{r_m r_p}
\]

Transmission torque from the motor gear, \( \tau_m \), to the base pinion gear, \( \tau_p \), can be modeled as in equation 8 where \( \eta_m \) 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
\[ \tau_m \geq \frac{r_m r_p \mu_s g}{n \eta_m \cos \theta} \left[ \sum_{i=2}^{n} (m_p(i) + m_r(i)) + m_p(1) \right] \] (9)

The cascaded linear actuator used in our proposed application consists of two tiers. The entire system is 3D printed using nylon material. From the material datasheet, the material onyx’s static friction coefficient, \( \mu_s \), 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, \( \eta_m \). Lastly, the summation of the racks and pinions masses, \( m_p(n) \) and \( m_r(n) \), is 0.167 kg, while the cable angle, \( \theta \), is 30 degrees at its max. Hence, using Equation 9, the minimum motor torque, \( \tau_m \), multiple with a torque safety factor of 2.25 is 0.01028 Nm, which is the minimum requirement to drive the cascaded linear actuator.

B. Pseudo-Stereo 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, \( b_{base} \).

Fig. 5: Overview of pseudo-stereo camera configuration featuring the custom cascaded bi-directional actuator.

Broadly, the overall camera assembly operation and our parallel pseudo-stereo camera setup calibration process remain the same as our previous work [2]. However, the translation from a standard single-tiered rack-and-pinion to an \( n-th \) tiered cascaded set-up changes the baseline calculation. In order to achieve desired overlap percentage \( O_{desired} \) between images for photogrammetry [12], 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

\[ b_{base} = \frac{W_x \times (1 - O_{desired})}{(2n - 1)} \] (10)

where \( W_x \) represents the horizontal camera FOV. The maximum horizontal FOV, \( W_{max} \), of the camera in this pseudo-stereo configuration is given as \( W_{max} = L_{max} + W_x \), where \( L_{max} \), which is the maximum effective travel length of the \( n-th \) 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 backlash that the cascaded linear actuator may face is taken into account. Even with the maximum backlash error of \( \pm 5\) 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 data set preparation, 3D 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 Network Pipeline

Training a highly accurate neural network model is dependent on the quality of the training data set in terms of its variance, quantity, and accurate representation of the target of interest. Obtaining such a data set is administratively expensive and challenging due to the target’s remote location. The wide range of different bearing series would mean a large data set is needed to cover all cases. Hence, instead of collecting on-site images to generate enough point clouds to form the whole training data set, 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 data set.

The data set generation pipeline involves first obtaining a point cloud model from an image set of the actual object, which is the 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 [13] monocular camera calibration process. The generation of point cloud is done using the python library of the Metashape software by Agisoft LLC1. The proprietary process of Agisoft uses SIFT algorithm [14] 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 Editor2. After generating all available real point cloud data, the rest of the point clouds that form the bulk of the training data set 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

\[ P_{ctr} = \{p_x, p_y, p_z \in \mathbb{R} | (\forall p_x - \mu_x), (\forall p_y - \mu_y), (\forall p_z - \mu_z) \} \] (11)

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

\[ P_{norm} = \{p_x, p_y, p_z \in \mathbb{R} | \frac{\sqrt{p_x^2 + p_y^2 + p_z^2}}{max_x(\sqrt{\sum_i |x_i|^2})} \} \] (12)

1https://www.agisoft.com/
2https://github.com/Hitachi-Automotive-And-Industry-Lab/semantic-segmentation-editor/
Finally, randomly select an augmentation method using a uniform distribution. The augmentation methods consist of six different types and it is described as
\[
\zeta_i \in [\zeta_{lb}^i, \zeta_{ub}^i], i = \text{Prob}(1|2|...|N)
\]
(13)
where \( \zeta \) is the augmentation method, and the strength of the augmentation is within and lower, \( lb \), and upper, \( ub \), bound, 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 random selected. The training data set 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. 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++ [15] is a hierarchy network architecture whose concept runs PointNet [16] recursively and progressively through an increasing scale. Its key differentiator from other 3D semantic segmentation network architecture is that it consumes labeled point cloud data directly as input, which helps to overcome irregular point cloud format issues. The network architecture employed consists of a 4-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 is shown in Table 1.

### Table 1: Model Training Results

<table>
<thead>
<tr>
<th>Loss</th>
<th>mIOU</th>
<th>Acc</th>
<th>Recall</th>
<th>Precision</th>
<th>F1 Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train</td>
<td>0.0499</td>
<td>0.9106</td>
<td>0.9362</td>
<td>0.8622</td>
<td>0.8711</td>
</tr>
<tr>
<td>Validate</td>
<td>0.0994</td>
<td>0.8520</td>
<td>0.8688</td>
<td>0.8622</td>
<td>0.8711</td>
</tr>
</tbody>
</table>

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 semi-circle 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 3D problem into a 2D space, using only the x and z axes points of the point cloud, \( P_{xy} = \{x, y \in \mathbb{R}\} \), 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 semi-circle 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, \( dbSCAN \), 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 [17] 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 [18], the hyper fit technique, which is translation and rotation invariant, yields the best result by combining two other methods Pratt [19] and Taubin [20], to eliminate essential bias. This result in best-fitted circles, \( O_{(b)}(ct_{(x)}, ct_{(z)}, rad) \), where \( ct_{(x)} \) and \( ct_{(z)} \) are the \( x \), \( z \) coordinate of the fitted circles’ center and \( rad \) is its radius.

Then, the length of the consolidated reference region points \( ref_i \), in the scale of a unit circle, is calculated by finding the distance between the points of edges using

\[
ref_i = \sqrt{(dx_{ref})^2 + (dz_{ref})^2}
\]

where \( dx_{ref} \) and \( dz_{ref} \), which is equal to \( \max(ref_{(x,z)}) - \min(ref_{(x,z)}) \), 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 semi-circle 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, \( \lambda \), 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

\[
a = dx^2 + dz^2 = (pr_{x} - pt_{x})^2 + (pr_{z} - pt_{z})^2 \quad (15)
\]

\[
b = 2[dx(pt_{x} - ct_{(b)\{x\}}) + dy(pt_{z} - ct_{(b)\{z\}})] \quad (16)
\]

\[
c = (pt_{x} - ct_{(b)\{x\}})^2 + (pt_{z} - ct_{(b)\{z\}})^2 - rad_{(b)}^2 \quad (17)
\]

where, \( a \), \( b \), and \( c \) are the variables. The left and right edge coordinates of the line projection, \( pt_{(x,z)} \) and \( pr_{(x,z)} \) are equal to \( (-1) [m(-1 - ct_{(b)\{x\}}) + ct_{(b)\{z\}}] \) and \( (1) [m(1 - ct_{(b)\{x\}}) + ct_{(b)\{z\}}] \), respectively. The intersection points of the best-fit circles’ left and right sides for the base and top plate and the line projection, \( lhs_{(b)\{t\}} \) and \( rhs_{(b)\{t\}}, \) is

\[
(lhs_{(b)\{t\}}x, lhs_{(b)\{t\}}z) = (dx \times rt_1 + pt_{x}, dz \times rt_1 + pt_{z}) \quad (18)
\]

\[
(rhs_{(b)\{t\}}x, rhs_{(b)\{t\}}z) = (dx \times rt_2 + pt_{x}, dz \times rt_2 + pt_{z}) \quad (19)
\]

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

\[
X_1 = \left[ \sqrt{(lhs_t - lhs_b)^2 + (rhs_t - rhs_b)^2}/ref_1 \right] (ref_a) \quad (20)
\]

\[
X_2 = \left[ \sqrt{(lhs_t - rhs_b)^2 + (rhs_t - lhs_b)^2}/ref_1 \right] (ref_a) \quad (21)
\]

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 \( X_1 - X_2 \). 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 paper. The actual dimension of the reference region is available from the bearing schematic. Most bearing scale bar’s reading fades off over time, making it unusable for the current inspection process, but our algorithm does not depend on the reading.

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

**IV. EXPERIMENTS AND RESULTS**

A. Unmanned Aerial Platform Description

The dimension of the custom BINOv2 UAV platform currently used for the proposed application is 265mm by 76mm by 200mm. The camera assembly’s maximum length, \( L_{max} \), which features a two-tier cascaded linear actuator, is 510mm. BINOv2 is slightly wider than its predecessor, where it was 231mm. However, BINOv1 only had a maximum actuator length of 177.53mm compared to the current version, 510mm length, while only sacrificing 30mm in its overall length. The viaduct bearing entrance’s overall size is 850mm by 150mm by 300mm, 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. The \( L_{max} \) with the camera renders an entire horizontal camera FOV, \( W_{max} \), of 694mm. Figure 8 and 9 shows the component breakdown of BINOv2 and the cascaded linear camera assembly, respectively.

**Fig. 8:** Components breakdown of the custom BINOv2 UAV.

B. Mock-up Viaduct Experiment

The remote and inaccessible location of the bearing prevents various controlled experiments that require human access.
Fig. 9: Components breakdown of cascaded bi-directional linear actuator camera assembly design.

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 2m. 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.

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

1) **Full Conduct Inspection Operation**: The first part involves conducting the full inspection operation to simulate 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 2m, 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 results for the third experiment. The results show that measurements obtained through a traditional 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.

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, but regardless, the algorithm still needs to produce accurate measurements. The second experiment consists of ten execution of the algorithm for each 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 0.822mm, which is still much within the current standards of ±3mm.

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 2D measurement algorithm from our previous work [2] 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 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.

<table>
<thead>
<tr>
<th>Displacement Measurement (mm)</th>
<th>10°</th>
<th>20°</th>
<th>30°</th>
<th>330°</th>
<th>340°</th>
<th>350°</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Displacement Type</strong></td>
<td><strong>Mean</strong></td>
<td><strong>Standard Deviation</strong></td>
<td><strong>Mean</strong></td>
<td><strong>Standard Deviation</strong></td>
<td><strong>Mean</strong></td>
<td><strong>Standard Deviation</strong></td>
</tr>
<tr>
<td><strong>Ground Truth</strong></td>
<td>6.000</td>
<td>0.546</td>
<td>6.354</td>
<td>0.504</td>
<td>0.757</td>
<td>0.892</td>
</tr>
<tr>
<td><strong>RMSE</strong></td>
<td>0.797</td>
<td>0.757</td>
<td>0.635</td>
<td>0.801</td>
<td>0.822</td>
<td>0.539</td>
</tr>
<tr>
<td><strong>Max Error</strong></td>
<td>1.233</td>
<td>0.989</td>
<td>1.168</td>
<td>1.212</td>
<td>1.235</td>
<td>1.104</td>
</tr>
</tbody>
</table>

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

**TABLE V: Proposed Algorithm vs Measurements from Commercial Stereo Cameras**

<table>
<thead>
<tr>
<th>Camera Type</th>
<th>Minimum Distance</th>
<th>Monocular</th>
<th>Intel RealSense 43Si</th>
<th>Structure Core</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>70mm</td>
<td>250mm</td>
<td>390mm</td>
<td></td>
</tr>
<tr>
<td>Ground Truth</td>
<td>6.000</td>
<td>6.000</td>
<td>6.000</td>
<td></td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.538</td>
<td>1.011</td>
<td>1.043</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>6.380</td>
<td>1.491</td>
<td>-0.155</td>
<td></td>
</tr>
<tr>
<td>RMSE</td>
<td>0.601</td>
<td>4.509</td>
<td>6.155</td>
<td></td>
</tr>
<tr>
<td>Max Error</td>
<td>1.012</td>
<td>5.786</td>
<td>8.022</td>
<td></td>
</tr>
</tbody>
</table>

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 8m 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, while the readings on the scale bar pre-installed 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 3D 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 1mm, which is better than the current accuracy standard of ±3mm.

**TABLE VI: Actual Site Trial Experimental Results**

<table>
<thead>
<tr>
<th>First Site Experiment</th>
<th>Ground Truth</th>
<th>Standard Deviation</th>
<th>Mean RMSE</th>
<th>Max Error</th>
<th>1.520</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.000</td>
<td>0.867</td>
<td>4.290</td>
<td>0.925</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Second Site Experiment</td>
<td>Ground Truth</td>
<td>Standard Deviation</td>
<td>Mean RMSE</td>
<td>Max Error</td>
<td>1.554</td>
</tr>
<tr>
<td>0.500</td>
<td>0.602</td>
<td>0.264</td>
<td>0.974</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Third Site Experiment</td>
<td>Ground Truth</td>
<td>Standard Deviation</td>
<td>Mean RMSE</td>
<td>Max Error</td>
<td>1.498</td>
</tr>
<tr>
<td>2.000</td>
<td>0.840</td>
<td>2.180</td>
<td>0.644</td>
<td>1.498</td>
<td></td>
</tr>
</tbody>
</table>

![Fig. 11: BINOv2 within the actual site cavity to capture stereoscopic images of bearing.](image)

V. DISCUSSION AND 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.

REFERENCES


