CAAP Quarterly Report

Date of Report: *October 7th , 2019*

Contract Number: 693JK31850007CAAP

Prepared for: *Robert Smith, Project Manager, PHMSA/DOT*

Project Title: *A novel structured light based sensing and probabilistic diagnostic technique for pipe internal corrosion detection and localization*

Prepared by: *Mohand Alzuhiri, Rahul Rathnakumar, Dr. Yiming Deng, Dr. Yongming Liu*
Contact Information: *Dr. Yiming Deng (MSU) and Dr. Yongming Liu (ASU)*

For quarterly period ending: *October 7th , 2019*

Business and Activity Section

(a) Generated Commitments

Project abstract: Internal corrosion in pipes is dangerous due to multiple factors contributing to its development. Degradation of pipeline health is susceptible to hazard due to failure. To prevent such failures, a major challenge for the maintenance crew to detect and repair corrosion still prevails due to difficult and expensive accessibility during scheduled maintenance. The proposed method will focus on the development of novel structural light-based imaging for internal corrosion detection, which simplifies the detection process while achieving superior spatial resolution. The proposed approach will develop an endoscopic structured light scanning tool that is based on phase measurement profilometry (PMP). The developed system will be simple to fabricate and easy to be used by maintenance personnel with minimal skillset due to its intuitive scans. The structured light system will be developed to generate high-resolution reconstructed images representing surface texture with high accuracy. Based on the images, additional processing capabilities developed using Bayesian updating technique will give the capability of automatic classification and identification of different types of precursors. A convolutional neural network-based corrosion detection method will provide automated detection, which further minimizes the operator involvement. The uncertainty quantification technique will be integrated to enhance the probability of detection and to quantitatively determine the damage size and location.

Based on the identified challenges and state-of-the-art techniques, a complete solution is needed to detect, localize and evaluate internal corrosion in metal pipelines. Our proposed solution is to develop a phase measurement profilometry (PMP) structured light-based tool to detect and locate any surface defects and damages on the pipe wall, which includes specifically corrosion. Internal corrosion will be effectively detected and evaluated by checking the amount of material loss, color change, spread and

pattern in the pipe wall simultaneously with a capability of integrating with ILI platforms. This optical inspection is also integrated with a set of numerical tools to evaluate structured light sensor information in order to automate the analysis of inspection data. The specific technical objectives/goals of the proposed research are:

- Design and develop a PMP structured light-based in-line inspection endoscopic scanner. The deliverables include:
    - design a new SL module to produce patterns with high resolution and contrast.
    - develop a new scheme to calibrate the new PMP based projector and camera(s).
- Develop a new reconstruction algorithm called moving phase measurement profilometry (MPMP) to exploit the system movement of the scanner along the pipe to enhance the quality of the reconstruction and detection.
- Evaluate the suitability of different optical methods like stereo cameras to enhance the performance of corrosion detection.
- Develop a convolutional neural network-based model for the automatic detection and classification of corrosion damages from the provided structured light sensor data to mitigate the need for manual analysis of 3D sensor massive data.
- Develop an uncertainty quantification technique to enhance the probability of detection and to quantitatively determine the damage size and location.

Educational Objectives: Another major objective of the proposed effort is to inspire, educate and train Ph.D. and MS students to address pipeline safety challenges, potentially as a career after their graduation. If funded, two Ph.D. students from both universities and several MS/undergraduate students will be included in this CAAP program. They will be trained and educated in science and engineering to address pipeline safety and integrity challenges. The PIs believe education is a critical component of the CAAP project, and we will integrate research with educational activities to prepare the next generation scientists and engineers for the gas and pipeline industry. Specific educational objectives include:

- Inspiring, educating and training the graduate students at MSU and ASU as research assistants for pipe integrity assessment and management. Our previous successful CAAP projects have produced several engineers, researchers and summer intern in gas and pipeline industry,
- Integrating research topics from this effort with the existing undergraduate research programs at MSU, e.g. ENSURE program at the College of Engineering and ASU to involve undergraduate students in pipe safety research.
- Improving the curriculum at MSU (e.g., Nondestructive Evaluation) and ASU (e.g., Machine Learning and Artificial Intelligence) using the scientific findings and achievement from the proposed research,
- Adapt research topics from this project to student projects in seminar, senior design, and project courses, in order to make educational impacts on broader groups of students,
- Encourage the graduate research assistants involved in this project and students in the courses to apply for internships at USDOT/PHMSA and industry to practice their learned skills and gain practical experiences in areas related to pipe safety and integrity.

The above-mentioned goals and objectives of this CAAP project will be well addressed and supported by the proposed research tasks. Development, demonstrations and potential standardization to ensure the integrity of pipeline facilities will be carried out with the collaborative effort among two different universities and our industry partner, Gas Technology Institution. This MSU-ASU-GTI team has successfully completed several PHMSA projects including "Slow Crack Growth" study, which was ranked No. 2 overall in all core PHMSA projects in 2017. The quality of the research results will be overseen by the PIs and DOT program manager and submitted to high-profile and peer-reviewed journals and leading conferences. The proposed collaborative work provides an excellent environment for the integration of research and education as well as tremendous opportunities for two universities supported by this DOT CAAP funding mechanism. The graduate students supported by this CAAP research will be heavily exposed to ILI, NDE, reliability and engineering design topics for emerging pipeline R&D technologies. The PIs have been actively encouraging students to participate in past and ongoing DOT projects and presented papers at national and international conferences. Students who are not directly participating in the CAAP project will also benefit from the research findings through the undergraduate and graduate courses taught by the PIs and attending university-wide research symposium and workshop.

(b) Status Update of Past Quarter Activities

Task 1 – Structured Light System Development

This project aims to provide a fast, robust and easy to use tool to detect and characterize corrosion and corrosion-related damage. In this task, we are aiming to fabricate and miniaturize a sensor that can be inserted inside gas pipelines and building a simple interface to guarantee its ease of use. The following sections will show the progress that MSU made in creating a new sensor that can be inserted inside a 6-inch pipe and also shows some initial images from the new setup.

In the past quarter, we reported that the scanning of a clean iron samples resulted in strong surface reflections that saturated the imaging sensor. The strong reflections in the system were related to the highly reflective surface of the clean iron samples. We also reported that the scanning system suffered from strong vignetting on the edge of the acquired image. This phenomenon can be related to the narrow field of view of the projection system and the strong vignetting from the projection lens. In the past quarter, we worked on solving the aforementioned issues of the rectangular system and also worked on the development of a new cylindrical scanning system for the pipe environment.

Creation of corroded samples:

MSU has created two corroded samples for the evaluation of the rectangular scanning system in order to increase the speed of sensor development. The samples are made of cast iron with machined surface defects that have different shapes and depths. A corrosive solution was created in order to have a fast reaction with the iron surface. The solution was created by using 6 milliliters of 1% Hydrogen peroxide (H2O2), 1 milliliter of acetic acid (Ch3COOH) and 2 grams of sodium chloride (NaCl). The samples were cleaned first to remove greases and dust and then sprayed with the solution and left in the open air for 30 minutes to create a thin layer of rust. The created corroded samples are shown in Figure 1. These samples provide a more realistic testing environment and have lower surface reflectivity.

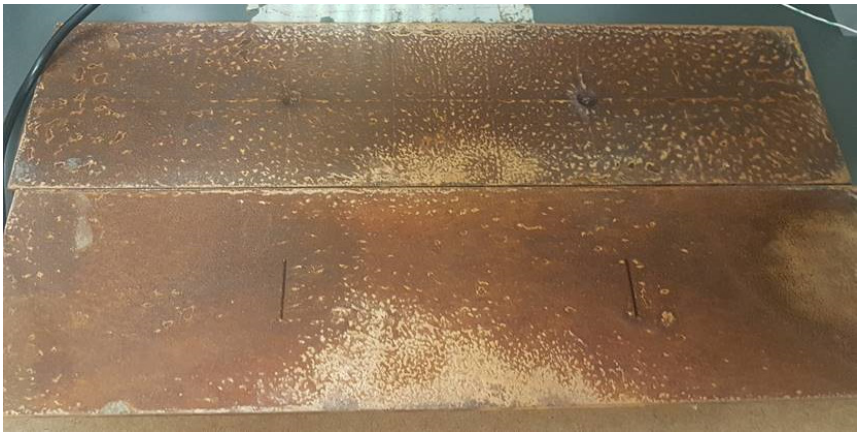


Figure 1: Corroded samples

Rectangular geometry sensor:

The first version of the rectangular geometry sensor had narrow projection angle which resulted in incomplete coverage of the camera frame. In order to reduce this effect, the lens was exchanged with a 120 degrees low distortion lens as shown in Figure 2. Replacing the lens resulted in increasing the size of the illuminated area in front of the camera and less vignetting near the edges of the frame as shown in Figure 3. The addition of the new lens is also associated with a reduction in the spatial frequency of the projected pattern because the pattern spread over larger areas.

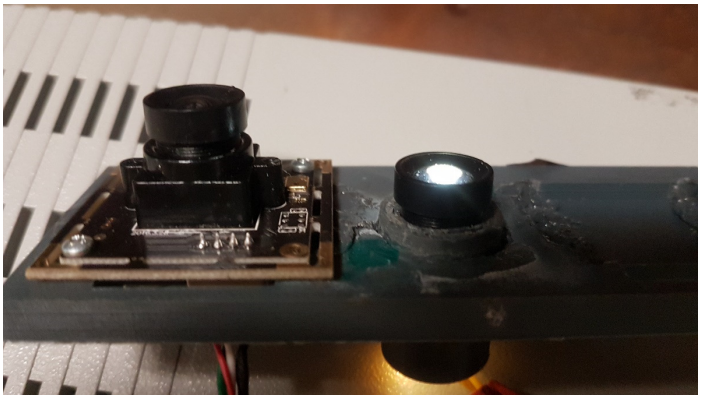
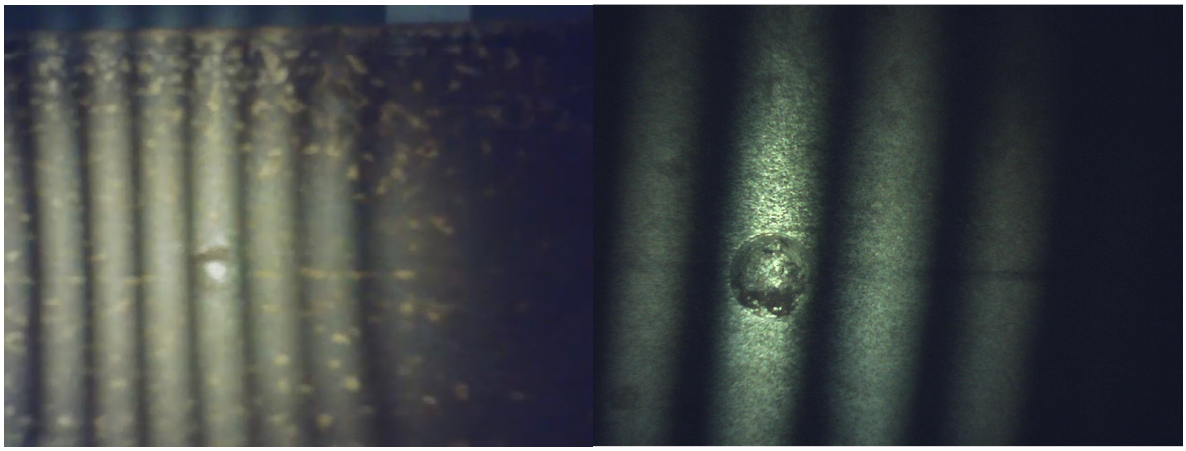


Figure 2: Rectangular scanner with the new lens



(a) (b)

Figure 3: a) System image with the old lens, b) System image with the new projection lens

Sensor design for cylindrical geometry:

Several design schematics were considered for the cylindrical system design inside the pipe environment. The proposed system schematics are shown in Figure 4. The schematic in Figure 4a consists of a camera and projector pointing in the same direction. The projector is chosen to have a relatively narrow field of view (76 degrees in our design) in order to project the image in front of the camera while the camera is chosen to have a much wider field of view (170 degrees in our design). This schematic was the main proposed design for the system but since this project is aimed for large diameter pipes, MSU opted to go with a more flexible design to enhance the system capabilities. The schematic in Figure 4b uses two stereo cameras that are facing the pipe wall accompanied by a projector that is aiming in the same direction. This design achieves 360-degree view by using multiple rectangular stereo systems to cover all the different directions. The main challenge with this system is that using multiple projectors causes interference of the projected pattern at the edges of the frames. To solve this problem, we opted to go with design C. In this design, multiple cameras are covering different parts of the field of view while a single projector is used to illuminate the scene. The multi-camera setup increases the resolution of the acquisition system and reduces the effect of image degradation from using very wide field of view cameras. The complexity of the design is also reduced by only a single light projector. Reducing the resolution of the projection system is not critical since we are already using lens defocusing in order to produce the needed sinusoidal fringes.

For the design in Figure 4a, MSU is planning to use the same prototype we designed for the colored multi-ring inspection system shown in Figure 5. The system consists of a SONY IMX219 CMOS sensor with 170 degrees camera and a 76 degrees projection module. The plan is to exchange the colored circular pattern with a new pattern with black and white circles. For design B, the same rectangular geometry scanner can be employed but after mounting it to a rotary platform in order to obtain 360 degrees field of view.

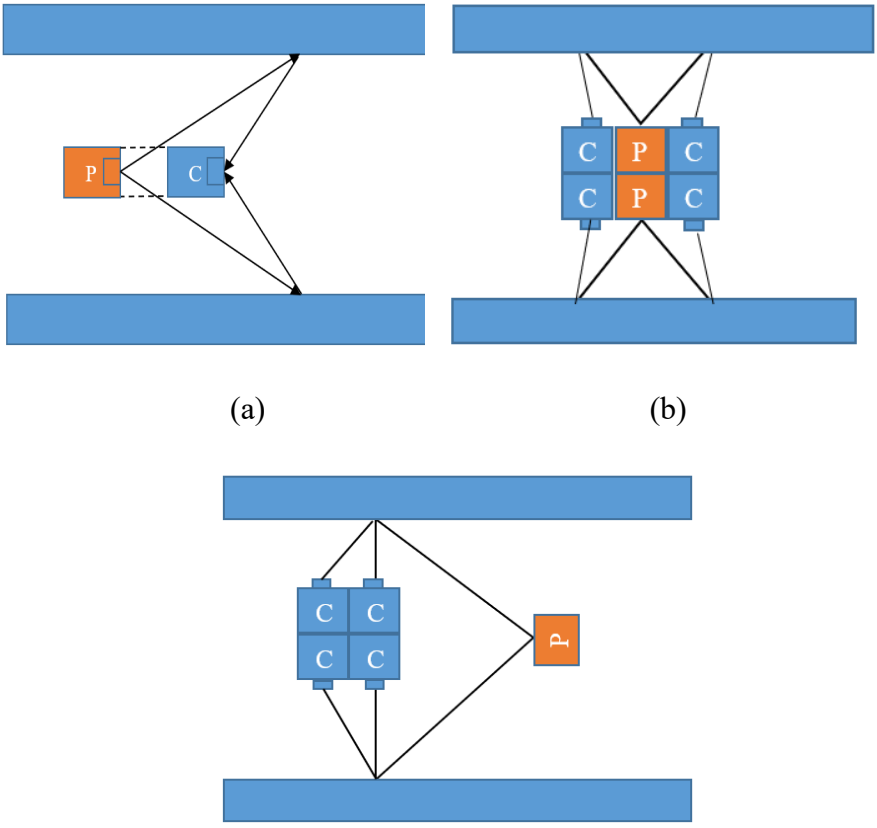
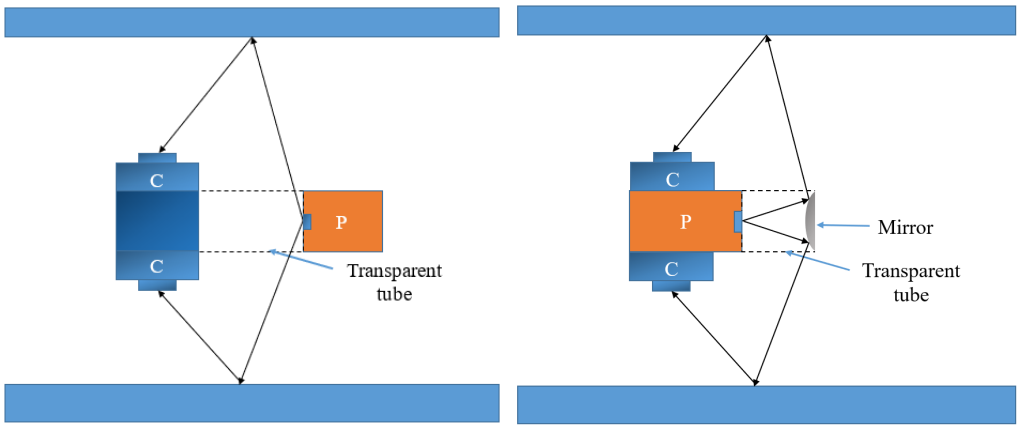


Figure 4: Proposed cylindrical systems

Figure 5: Endoscopic structured light sensor

Currently, most of the work is devoted toward developing a prototype for model C. Direct implementation of the model is explained in Figure 6a. One of the drawbacks of this model is the wire connection needed to power the slide projector and the limited depth of field of the projection lens. The problem of the connections can be solved by coating the power connections on the surface of the transparent glass tube by using plasma sputtering. The depth of field can also be extended by closing

down the lens aperture, but this will also result in reducing the brightness of the projector. Another solution is based on using a projector coupled with a mirror in order to create a backward-looking projector as shown in Figure 6b.



(a) (b)

Figure 6: Proposed implementation of model C with a) Direct implementation, b) Backward looking projector with a spherical mirror

This type of projectors offers the following advantages:

- Eliminate the need for the sputtering process needed to power the projector
- Reduce the total size of the projector since the projector hardware can be embedded beneath the stereo cameras.

On the other hand, the system has some drawbacks that can be summarized as follows:

- It requires high precision during the fabrication process to guarantee the correct alignment between the main projector and the reflective mirror.
- Requires a high-quality mirror to increase the system efficiency and have bright projected pattern on the pipe wall.

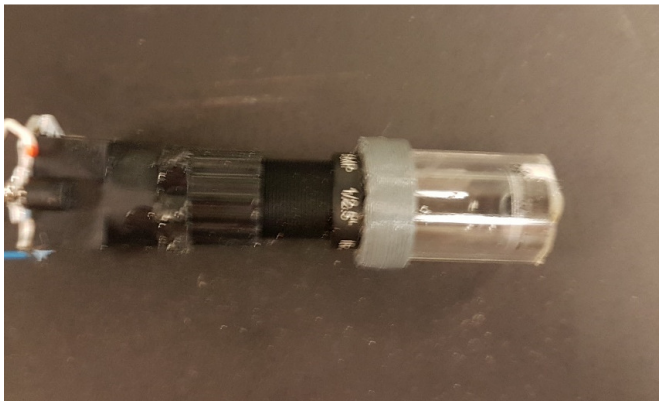


Figure 7: Fabricated projector with a spherical mirror

An initial prototype of the backward-looking projector was created as shown in Figure 7. The main projector was created by using a high-power LED and a black and white slide with concentric rings

followed by a 20 degrees projection lens. The main projector is coupled with 12mm Dia x -6.1mm FL Enhanced Aluminum Coated, Convex Mirror. The projector was fabricated and tested successfully but further improvements are needed to enhance the projected image quality.

Task 2 –Reconstruction of corrosion damage

In this section, we discuss the most recent results regarding the surface reconstruction results of the scanned reference samples. The current results were collected with the uncalibrated rectangular scanning system; therefore, the results are prone to some periodic errors due to uncorrected vignetting and lens distortion. The results also represent relative height maps that have no metric units.

Scanning with the initial scanning system:

In the last quarter, we reported that scanning uncorroded sheets of cast iron resulted in a large amount of specular reflection that saturated the imaging sensor. In order to have a reference sample that with small surface reflectivity, a sheet of cardboard was used as a scanning target. A shallow defect (Dent) was introduced to the surface of the scanned object as shown in Figure 8. The object was scanned with the initial prototype (with 76 degrees lens) and four images were collected as shown in Figure 9. The four images were registered, and a 3D map of the scanned was created as shown in Figure 10. Figure 10 represents a negative 3D map of the surface where we can clearly identify the location of the defect as marked by the red box.

Figure 8: Scanned cardboard sample

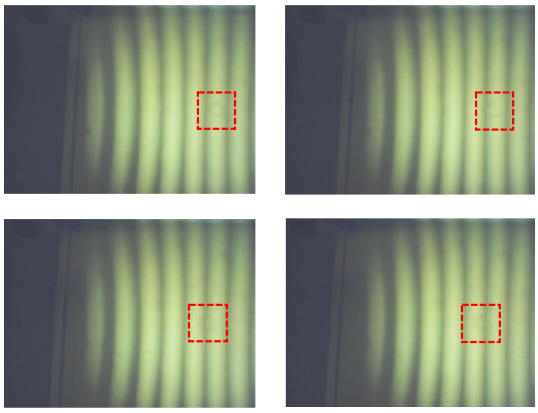


Figure 9: Acquired imaging sequence for cardboard sample

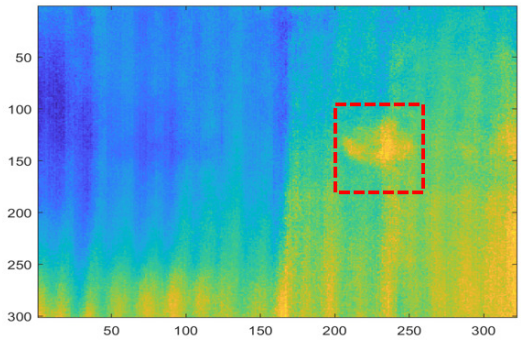


Figure 10: Reconstructed 2D height map for cardboard sample

Scanning with the updated scanning system:
The updated scanning system with the new lens was used to scan the newly prepared corroded sample. Each sample was scanned, and four images were acquired from each scanning process. The speed of the scanner was estimated by monitoring a reference known object under the scanner. The scanner was estimated to need 24 frames to move 360 degrees phase shift. The spatial frequency on the scanned surface was estimated by measuring the distance (in pixel) between every two minima or maxima as explained in Figure 11. After obtaining the system speed and spatial frequency of the pattern, the images were registered, and the height maps were calculated according to the following equation:

$$\phi(x,y) = \text{atan}^{-1}\left(\frac{I_2 - I_4}{I_1 - I_3}\right),$$

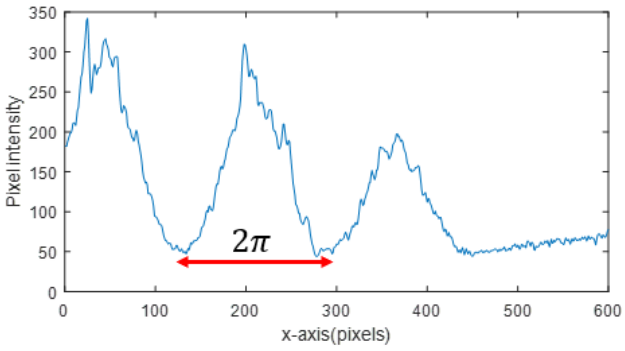


Figure 11: Calculation of the spatial frequency

In addition to the height maps, tree other components from the images were calculated which can serve as contrast-enhanced 2D images to enhance the probability of the damage detection. The first component is the original defect image under ambient light (DC value). The DC value can be calculated by summing the values of all the registered images.

$$DC_{image} = (I_1 + I_2 + I_3 + I_4)$$

This view provides a high contrast view of the scanned object without considering the effect of the sinusoid fringes. Another component that can be extracted is the intensity of the modulation sinusoid signal (AC). This component can be calculated by solving the original set of linear equations.

$$AC_{image} = \sqrt{(I3 - I1)^2 + (I_4 - I_2)^2}$$

The last component that can be extracted is the modulation ratio between the AC and DC values (γ).

$$\gamma = \frac{AC_{image}}{DC_{imge}}$$

The following sections give details about the scanning operation and the extracted components extracted from each defect. The sections show different cases with different shapes and depths.

Circle1

This case presents a dome-like defect with a depth of 2.1mm that has been punched into the surface of the scanned sample. The acquired imaging sequence is shown in Figure 12. The reconstructed components from the scanning process are shown in Figure 13. The reconstruction results show that the DC and AC components are clearly showing the defect with enhanced contrast. This contrast is almost lost when it comes to the modulation ratio between the AC and DC components. The 3D reconstruction is also showing that the defect area is detected with a circular shape and lower height value.

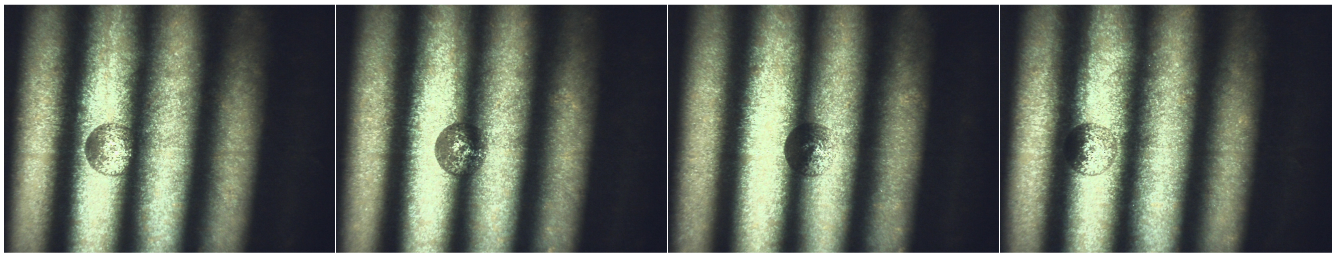


Figure 12: Image sequence for the second circle

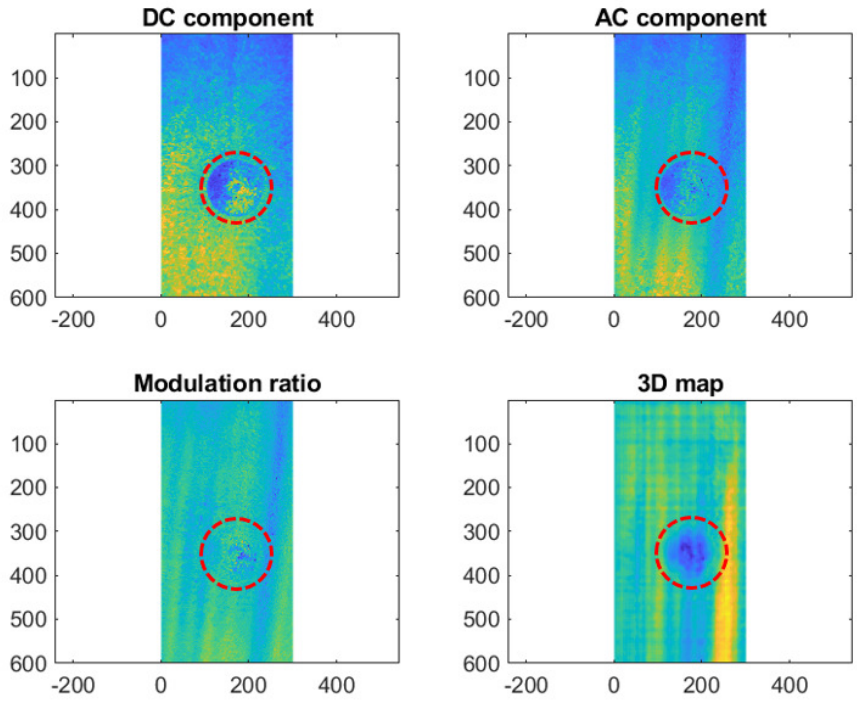


Figure 13: Reconstruction results for the first circle

Circle 2

This case presents a dome-like defect with a depth of 1.5mm that has been punched into the surface of the scanned sample. The acquired imaging sequence is shown in Figure 14. The reconstructed components from the scanning process are shown in Figure 15. The reconstruction results show that the DC and AC components are clearly showing the defect with enhanced contrast. This contrast is degraded when it comes to the modulation ratio between the AC and DC components. The 3D reconstruction is also showing that the defect area has lower height value which is consistent with the ground truth from the scanned sample.

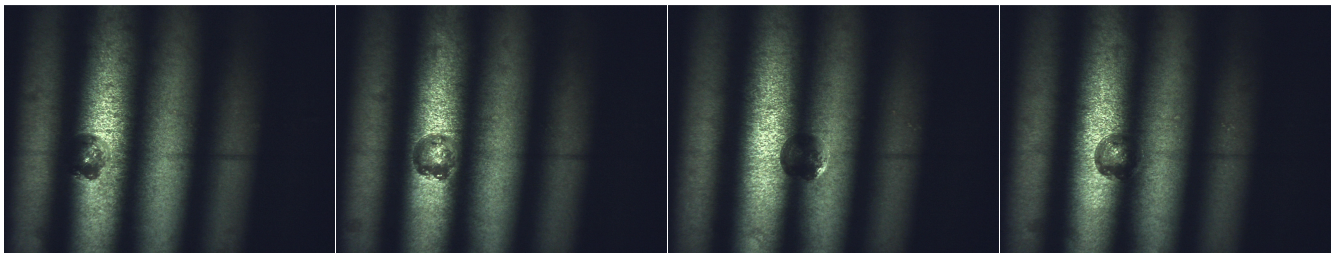


Figure 14: Image sequence for the second circle

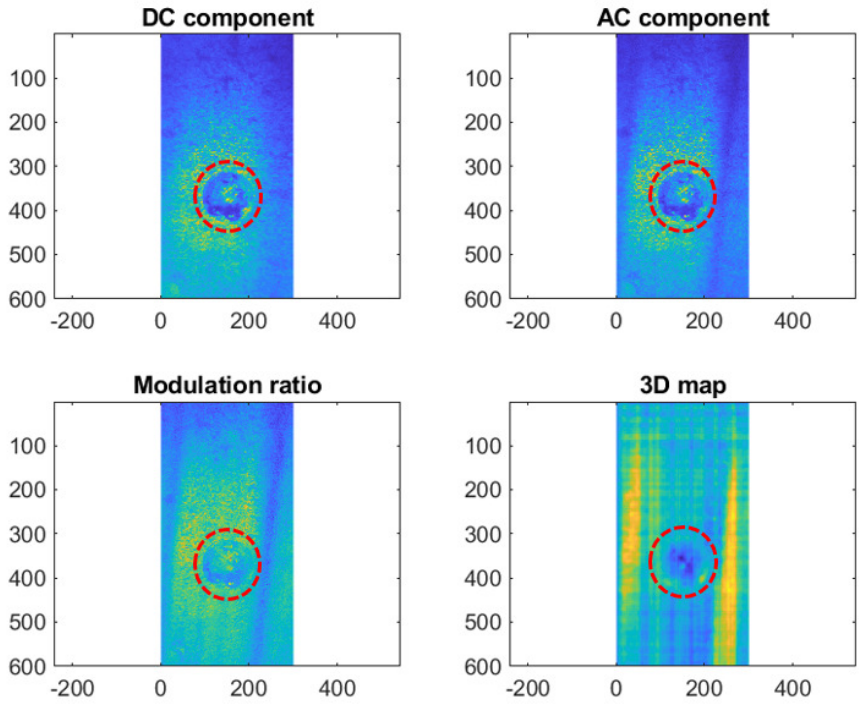


Figure 15: Reconstruction results for the second circle

Slit 1

In this case, a rectangular defect was cut into the metal surface with a depth of 1.2 mm. The scanning imaging sequence is shown in Figure 16. The components reconstruction results are shown in Figure 17. In the figure, we notice that the DC component has the highest contrast between the three 2D components followed by the AC component and then the modulation ratio. The 3D reconstruction also shows that the algorithm was able to detect the loss of materials where we have a lower height value.

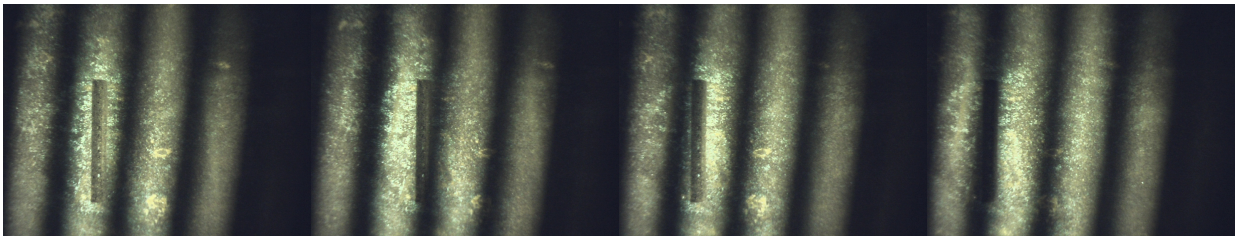


Figure 16: Image sequence for the first slit

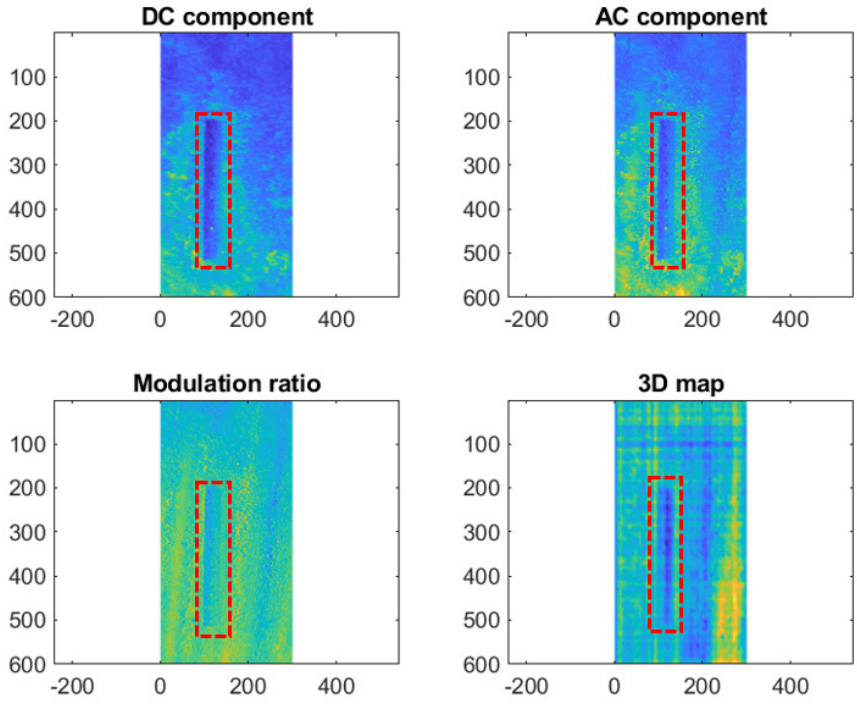


Figure 17: Reconstruction results for the first slit

Slit 2

This sample represents a rectangular defect with a depth of 0.7mm. The acquired image sequence is shown in Figure 18. The image analysis by using the three derived 2D components shows that the image contrast was improved and the slit can be clearly identified as shown in Figure 19. Careful examination of the 3D map inside the dotted red box shows that the shape of the slit was reconstructed but with poor contrast and the reconstructed object is barely noticeable from the background noise. This can be related to the shallow depth of the defect which makes it buried under noise from the system vignetting.

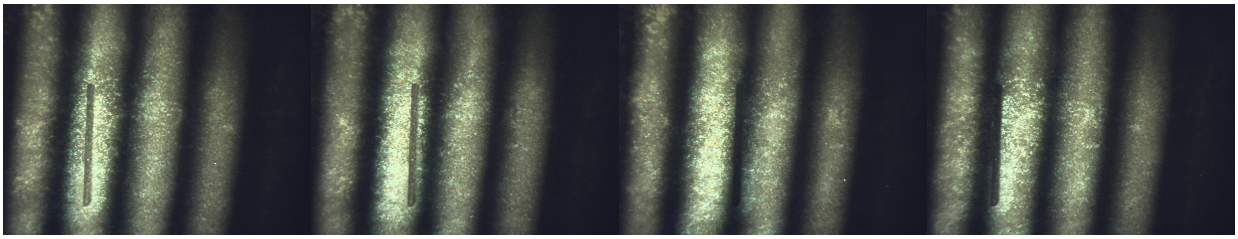


Figure 18: Image sequence for the second slit

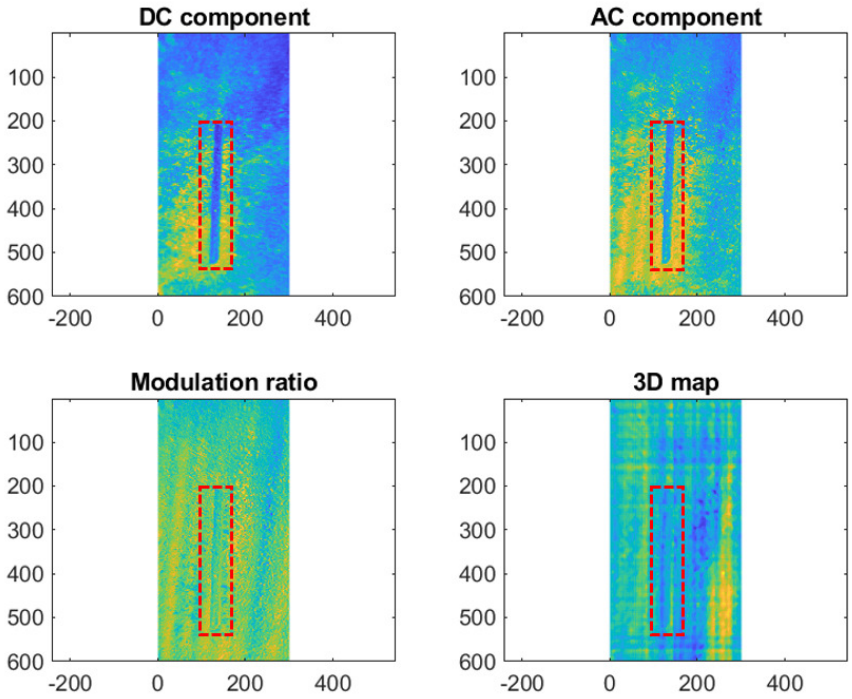


Figure 19: Reconstruction results for the second slit

Summary:

In this report, MSU focused on improving the performance of the scanning system by improving the quality of the hardware and also demonstrating the capability of the system to detect and reconstruct small size defects. The rectangular geometry scanning system was upgraded to provide a larger field of view and to enhance the illumination of the edges of the frame. Initial designs for the cylindrical scanning system were completed and the fabrication of a mirror-based back-projection scanning system was started. Two corroded samples were created by MSU to serve as a reference to evaluate the system performance. The two samples provided four defects with different shapes and depth into the metal surface. These defects were scanned by the upgraded scanning system and three 2D features were extracted from the scanning in addition to the 3D map. The 2D features provided a method to improve the contrast of the defects by combining the four scanning images. Uncalibrated 3D reconstruction was performed and the capability of the system to detect small size defects was demonstrated. Future work will be devoted to the fabrication of the cylindrical scanning system and finishing the system calibration.

Task 3. Automated corrosion detection and uncertainty quantification

1. Overview

The primary driver for Artificial Intelligence-based solutions for in-line inspection is the sheer size of piping infrastructure in refineries and cities. Full automation of the inspection procedure can reduce costs and drive higher reliability because automated inspection techniques are not susceptible to human error. The pipeline infrastructure is designed to last decades, and improper maintenance has significant safety and economic costs. The interactions between the fluids flowing through the pipe and the inner wall cause the degradation of the mechanical properties of the pipe over time, and early prognosis of damage can prevent catastrophic events.

The surface defect detection problem is the focus of this report. We introduce a new method of detecting cracks and compare it to the deep-learning method. We also investigate the shortcomings of the existing models and explore an existing method in literature to evaluate the performance of that model for our problem. Finally, we discuss the work to be done in the next quarter.

2. Previous work

The work done in the previous quarter focused on the evaluation of YOLO, a fully supervised method and the development of a novel, weakly-supervised deep learning method for crack detection. The deep learning algorithm developed in the previous quarter and the random forest algorithm are collectively called Crack Detection using Contour Features (CDCF). A fast method for analyzing image informativeness was also presented. Detailed metrics of the fully supervised YOLO algorithm resulted in good performance on data that wasn't far outside the domain of the test set. It also required bounding-box annotations of hundreds of images, which were hard to obtain. Therefore, the idea of learning from texture samples and testing on full images was explored, with promising results. Training on textures annotated only at the image-level generalized well to localize cracks on pipeline image data that had much more complexity, including variations in color, blurriness, and shape of cracks.

3. Random Forest Based Crack Detection Using Contour Features

1. Overview

Random forests are a method of building multiple uncorrelated decision trees, and then using a voting process to determine the class of the predictor variable for each testing example. The algorithm can be summarized as follows:

Training:

Consider a data-set with 1 prediction variable and k-predictors.

For trees t=1 to T:

1. Draw a sample X with size N from the training set.
2. Grow a decision tree by repeating the following steps for each terminal node until the minimum node size n_{min} is reached:

1. Select m variables at random from the p variables.
2. Pick the best variable/split-point among the m variables.
3. Split the node into two child nodes.

Prediction:

Let $C_b(x)$ be the class prediction of the b^{th} tree in the forest. Then $C_{rf}^{B}(x)$ is the majority vote of all the predictions from all trees $C_b(x)$.

The random forest algorithm needs a structured data-set for training, where we have a well-defined set of features for each image. The features were selected so as to be descriptors for the shape and textures of the crack regions. Shape and texture descriptors are powerful, condensed representations that encode information about the image. The Hu Invariant and the Grey-Level Co-Occurrence Matrix (GLCM) were used as shape descriptors for the problem. A brief description of these features are provided in the following section. The dataset used for training was the concrete crack classification dataset, a large collection of crack and non-crack textures, with 40,000 images. The test set was videos obtained from YouTube.

2. Methodology

i. *Shape descriptor: Hu Invariants*

Image moments are used to describe the distribution of pixels in an image. Hu invariants are moments that are invariant to rotation, scale and translation. This is a 7-dimensional feature vector that is obtained from image central moment calculations as follows:

$$\mu_{pq} = \iint (x - \overline{x})^p \, (y - \overline{y})^q f(x, y) dx dy \text{- (1)}$$

where p and q are the p^{th} and q^{th} order of the moment and $\overline{x}, \overline{y}$ are centroids. These central moments are invariant to translation. Scale invariance is introduced by normalization equations:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}}, \gamma = \frac{(p+q+2)}{2}, (p + q) = 2,3, \ldots \text{- (2)}$$

Based on these normalized central moments, Hu introduced 7 moment invariants that are invariant to rotation, that can be referenced in [1]. These moments are extracted from the edge map of the image patch, which gives us a discriminative representation of the crack and non-crack texture, as shown in figure upon which Hu invariants are calculated.

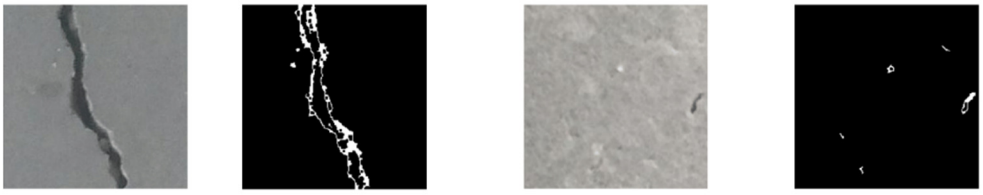


Illustration 1: Representative examples for Hu Invariant calculation: Edge maps make the invariants more discriminative

ii. *Texture descriptor: Grey-Level Co-Occurrence Matrix*

Gray level co-occurrence matrix (GLCM) is a simple texture descriptor that encodes the spatial relationship between pixels. This descriptor calculates how often a particular pixel combination occurs, with parameters being the spacing between the pixels and the orientation of the pixels. The distribution is given as:

For an offset $\delta x, \delta y$:

$$C_{\delta x,\delta y} = \sum_{x=1}^{n} \sum_{y=1}^{m} \begin{cases} 1 \\ 0 \end{cases} I(x,y) = i \wedge I(x+\delta x, y+\delta y) = j \text{ - (3)}$$

An 8-bit image would therefore have a 255x255 dimensional representation for one particular combination of orientation and spacing parameters. To reduce the dimensionality of the feature, and to cover a larger range of parameters, we encode the 8-bit image into a 4-bit image, as follows:

$I_{8\text{-bit}} = [0,63] \rightarrow I_{4\text{-bit}} = 0$ - (4a)

$I_{8\text{-bit}} = [64,127] \rightarrow I_{4\text{-bit}} = 1$ - (4b)

$I_{8\text{-bit}} = [128, 191] \rightarrow I_{4\text{- bit}} = 2$ - (4c)

$I_{8\text{-bit}} = [192,255] \rightarrow I_{4\text{-bit}} = 3$ - (4d)

This gave us a compact representation in the pixel space, resulting in a 16-dimensional feature vector for each combination of the orientation and spacing parameters.

iii. *Evaluating Splits in the Tree:*

Tree splitting quality is measured using the Gini Impurity criterion as the cost function. It is a measure of how often a randomly chosen element from a set would be incorrectly labeled if it was randomly labeled according to the distribution of labels in the dataset. A perfect value for the Gini impurity is 0. Gini impurity for a dataset with K-classes and each class has p_i items in it is:

$$G = 1 - \sum_{i=0}^{i=K} P_i^2 \text{ - (5)}$$

iv. Towards real-time performance: Agglomerative contour clustering

The contours obtained using the canny edge map were used as regions of interest for the detector. These were noisy, disjoint maps with high variance in contour area and length. Many disjoint contours were also neighbors, and the algorithm originally processed each of these contours individually. This reduced the detection speed of the algorithm. Here, we demonstrate a method by which nearby contours are clustered together in a bottom-up fashion, without losing detection performance and increasing detection speed by ~3x.

Agglomerative clustering methods have been used for discrete optimization problems and obtaining statistical metrics for numerical data. These methods do not require the number of clusters to be pre-specified, but the clusters emerge according to the metric being used to form them. In our case, we use the L2-distance between the contour centroid as our metric. The method is set up as follows:

$d \epsilon R^{SxS} where d_{ii} = 0 \wedge d_{ij} = d_{ji} \forall x, y \epsilon S$- (6)

d is a distance matrix. The parameter for this algorithm is the distance threshold, t, above which the algorithm does not combine contours.

Consider a set of S contours with centroids $[c_{xi},c_{yi}]$. We first create a symmetric distance matrix $d_{|S|x|S|}$ where $d_{ii} = 0$ and $d_{ij} = d_{ji}$ from the contour centroid set. The algoritm to combine contours works as follows:

Algorithm:

repeat until d(i,j) > $d_{threshold}$:

Determine a pair of mutually closest contours (a,b) from the distance matrix.

Join (a,b) to form new contour “ab”. Delete (a,b) from the set of nodes in the distance matrix and add “ab” to it.

Update the distance matrix with the new distances

The algorithm works at approximately $O(n^2)$ and with the number of contours usually less than 200, the frame rate is improved by ~3x due to this approach, delivering real-time speeds for the random forest-based detector.

3. Results and Discussion

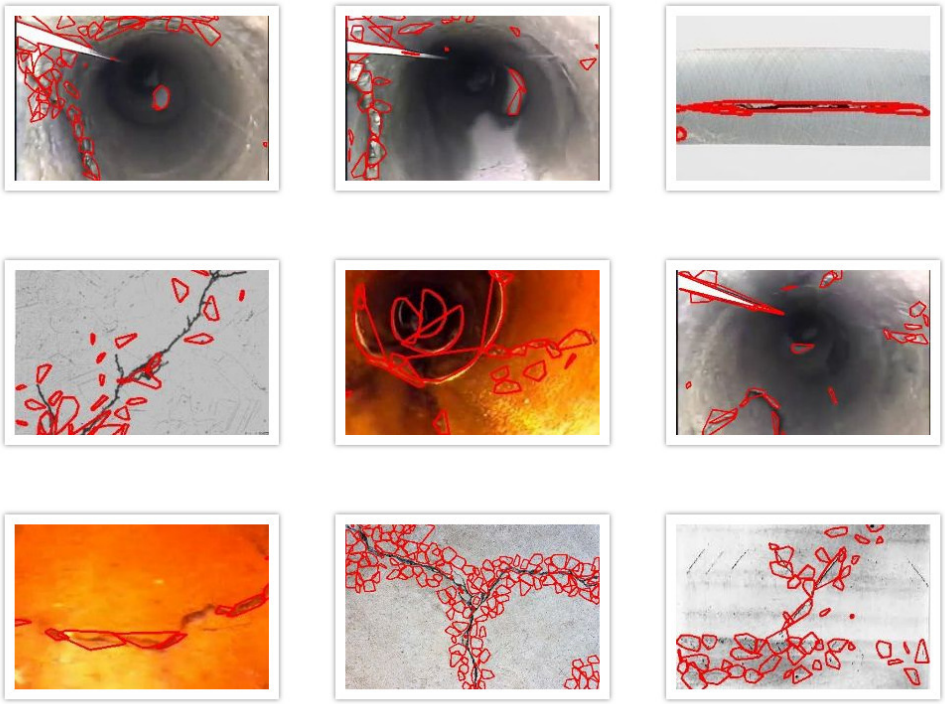


Illustration 2: Detection results from the random forest model

The random forest model was optimized using Randomized Grid Search. The parameters in the search space were:

S. No.	Parameter	Search Space Range
1	Number of Trees (n_estimators)	[10,100] Step size: 10
2	Number of Features to consider at each split-point (max_features)	[auto, sqrt]
3	Maximum number of levels in the tree (max_depth)	[10, 100] Step size: 10
4	Minimum number of samples to split a node (min_samples_split)	[2,5,10]
5	Method of selecting samples for training each tree (bootstrap)	[True, False]
6	Number of data folds in cross-validation	5
7	Number of combinations of parameters	ALL

Table 1 : Parameters space for randomized search

The search of all combinations of these parameters yielded the following results for various dataset sizes:

Dataset Size	Parameter values
1000	'n_estimators': 10, 'min_samples_split': 2, 'max_features': 'sqrt', 'max_depth': None, 'bootstrap': False
2000	'n_estimators': 20, 'min_samples_split': 2, 'max_features': 'sqrt', 'max_depth': None, 'bootstrap': False

Table 2 : Results of search in different dataset sizes

The number of estimators was restricted to 10, as further increases impacted detection performance without any measurable increase in performance metrics in the test dataset. The training dataset consisted of concrete crack dataset textures, that were all 60x60 texture patches, and the test set was video sources of pipeline inspection. For calculating performance metrics, 80 images from these video sources were selected in such a way that it captured a variety of backgrounds and crack shapes. Ground truth segmentations were created for these images, and the detections were compared with these segmentations. The metrics that were used to measure performance were the percentage of ground truth pixels covered by the detector, and the number of false positive detections. A false positive is defined as a detection whose IOU with the ground truth was less than 25%. The results for various dataset sizes are tabulated below:

Training set size	% Pixel coverage	Number of False Positive Detections per image
1000	52.38	26.38
2000	**61.66**	29.275
10000	57.98	26.575
40000	56.21	**25.1**

Table 3: Model performance with different training set sizes

Edge maps can often be noisy even in non-crack textures, reducing the discriminativeness of the Hu Invariant algorithm, and potentially reducing the precision of the model by increasing false positives. Suppressing this noise requires convolving a Gaussian smoothing kernel to the image patch. The following results tabulate the effect of including and exlcuding a (5x5) Gaussian smoothing kernel with a variance of 3 centered at 0.

Train set size	Smoothing	% Pixel coverage	Number of False Positive Detections per image
40000	None	46.89	24.6
40000	Gaussian (5x5) ~ N(0,3)	**56.21**	25.1

Table 4: Effect of Gaussian smoothing on the performance metrics of the model

4. **A Comparison between the deep learning and random forest method**

This section details how the deep-learning and the random forest method compare in terms of detection speed and model performance. The entire training dataset was trained on both the random forest and the deep neural network, and the detection pipeline produced the following results in the test set:

Method	Ave. Number of False Positives per Image
Deep Learning	5.65
Random Forest	23.53

Table 5: Number of false positives per image in the deep learning and random forest method

The deep learning method outperforms the random forest detector in both pixel coverage and false positives, however, it is an order of magnitude slower than the random forest.

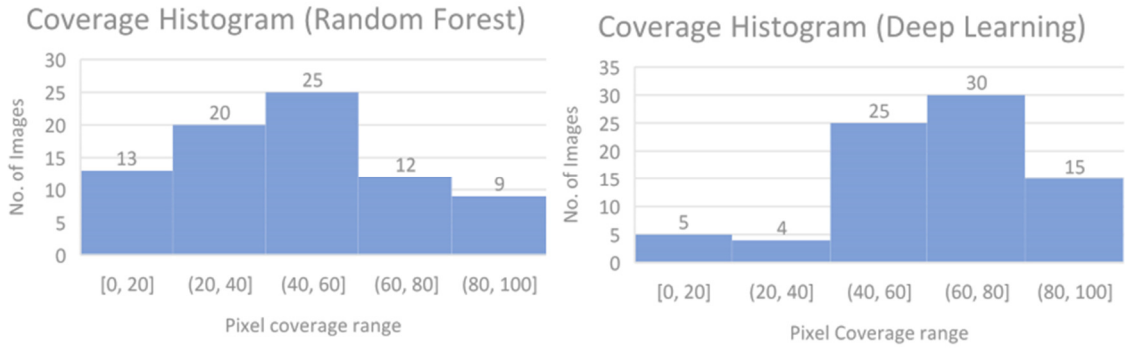


Illustration 3: Pixel coverage histograms for the test set for deep learning and random forest methods

5. Weakly Supervised Deep Detection Network

1. Overview

The current deep-learning algorithm performed weakly-supervised detection, and was able to detect cracks in pipeline videos after training with textural data. The speed of the architecture, however, was not real-time. Bilen Et. Al, [2], proposed a weakly-supervised deep detection network (WSDDN) that was explored for its performance and compared against our algorithm. While the Crack Detection Using Contour Features (CDCF) method collected the regions of interest and classified the raw patches via the CNN or random forest, the WSDDN works with the convolutional feature map of the image and regions of interest.

2. Architecture description

The architecture takes as input an image **x** and a set of regions of interest (RoIs) obtained through Selective Search [3] and outputs a set of bounding-boxes for the objects of interest. The architecture is shown in the figure below:

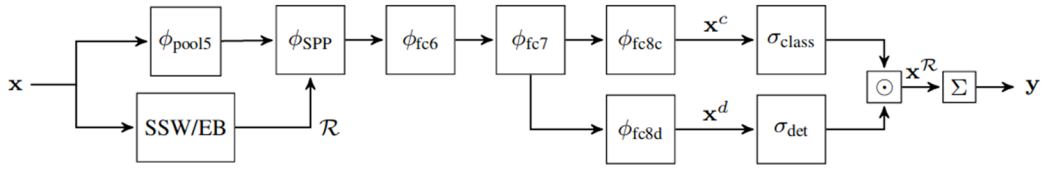


Illustration 4: WSDDN Architecture

ϕ_{pool} is the output of final convolutional layer from a classification network such as AlexNet or VGG19. ϕ_{SPP} is the RoI pooling layer. σ_{class} and σ_{det} are softmax outputs for the classification and detection data streams. These outputs are multiplied together and summed up to return an image-level classification score.

i. *Convolutional feature extraction:*

Convolutional feature extraction is accomplished using the VGG-19 or AlexNet backbones, where only the convolutional layers are utilized, and the fully connected layers are discarded and replaced with the downstream WSDDN layers.

ii. *Region of Interest pooling:*

The RoI pooling layer collects all the raw-image patches generated by Selective Search, and matches it to corresponding locations in the convolutional feature map from the previous section. This representation is max-pooled to fit the fully connected representation length.

iii. *Classification data stream:*

The regions of interest are pooled and passed through 2 dense layers, to yield a flat representation, from the matrix representation. This data is passed through to the classification data stream, where each invididual region is mapped to a C-dimensional vector of class scores, and then passed through a softmax operator to yield a C-dimensional vector proportional to the probability distribution.

$$[\sigma_{class}(x^c)]_{ij} = \frac{e^{x_{ij}^c}}{\sum_{k=1}^{C} e^{x_{kj}^c}} \text{ - (7)}$$

iv. *Detection data stream:*

The detection data stream runs in parallel with the classification data stream, and this stream computes for each individual class, the scores of different regions. A region that is more informative would have a high detection score.

$$[\sigma_{det}(x^c)]_{ij} = \frac{e^{x_{ij}^c}}{\sum_{k=1}^{R} e^{x_{ik}^c}} \text{- (8)}$$

v. Region-level scoring using Hadamard Product:

The region-level score for each region is then obtained by taking the hadamard product of (7) and (8) and are ordered based on this score. Non-max suppression is performed on these regions to remove regions with high overlap.

vi. Modification to the architecture:

Consider an image with n-regions of interest. Eqn (8) gives us the matrix of scores in the detection data stream, and these scores must add up to one across all regions for each class. Therefore, as n grows, the relative likelihood of a class in each region reduces. This would lead to lower total region scores. The threshold for region selection after this stage to pass into the non max suppression code becomes arbitrary. Therefore, the network was reduced by removing the detection data stream. The results from the classification data stream were operated on as follows:

$$C_{image} = \frac{\sum_{j=1}^{R}[\sigma_{class}]_{ij}}{N_{regions}} \forall i \epsilon C \text{ - (9)}$$

The required classes to be deteted can then be extracted from (7) and passed for non-max suppression.

3. Methodology

The experiements involved validation of the model on the benchmark dataset, followed by experiments on crack detection. The dataset used for benchmarking was the PASCAL VOC dataset, which contains 21-classes.

i. Training the network

The network is trained for our problem with each training image containing only one region of interest, ie, the entire image. The dimension of x^c for the classification data stream would therefore be 1x|C|. The loss function used for training is:

$$E(w) = \sum_{i=1}^{n}\sum_{k=1}^{C} log\left(y_{ki}\left(\phi_k^y\right)\right)$$

Stochastic Gradient Descent (SGD) with momentum is used to optimize the network.

4. Preliminary Results

The results pertaining to crack detection are presented here. The results did not yeild any improvements to the in-house methods that were developed. The localization performance was significantly worse, owing to selective search generating class-agnostic boxes and to the high recall of the model but low precision, as evidenced by the figure below.

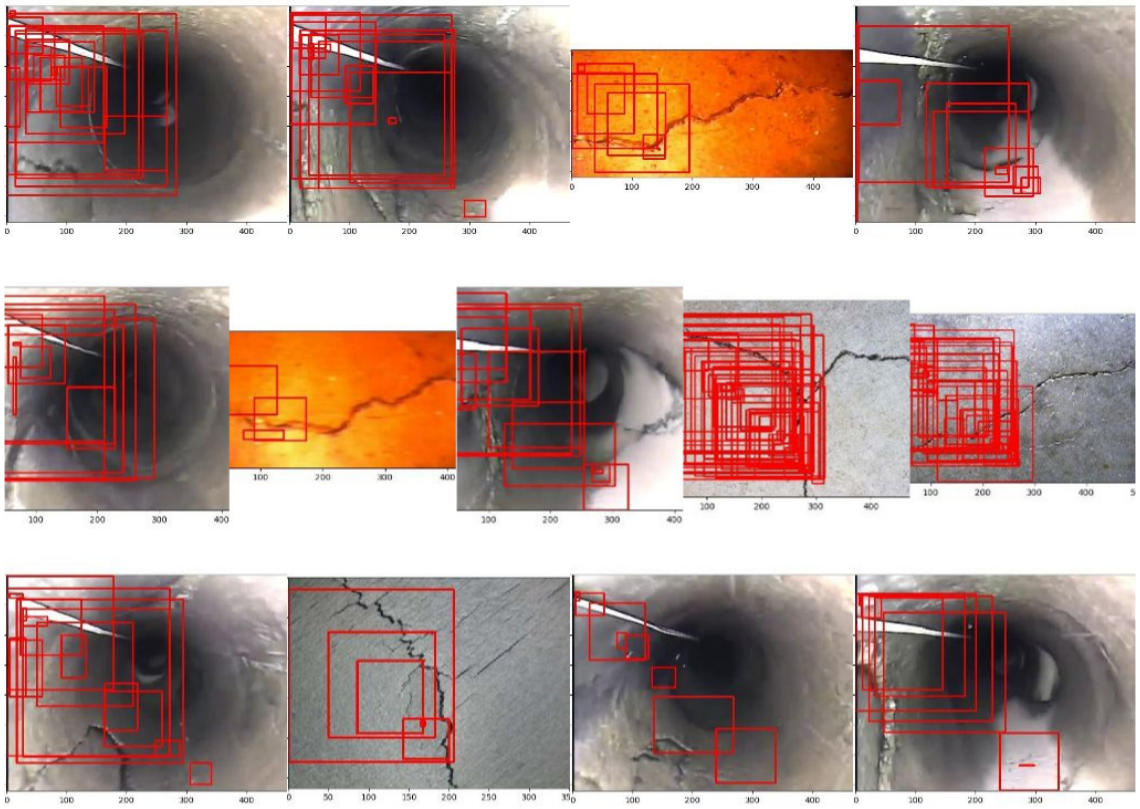


Illustration 4: Some detection results from the modified WSDDN algorithm.

Visual inspection of these results clearly demonstrate that the WSDDN is inferior in performance to CDCF algorithms that were developed earlier. The reasons for this is that the convolutional feature maps that the regions of interest get pooled into are highly compressed representations of the data. The discriminativeness of the data is therefore reduced compared to the deep learning CDCF method, where raw-image patches of the regions of interest are processed by the CNN. Increasing data-set size or training for a higher number of epochs does not remedy this drawback greatly.

6. Future Work

Given the results of the WSDDN, we have decided to seek more promising directions to improve the performance of the CDCF method. The crux of the problem in the CDCF method is that each region is processed serially in the CNN and finding ways to improve the speed of this processing step is key to achieving real-time performance. Another area we would investigate in the upcoming quarter is to expand the feature set of the random forest algorithm by introducing data-driven filter learning, as the random forest has already achieved real-time performance. Work in the next quarter would also focus on methods for region of interest detection for corrosion data. The good performance of the CDCF method can in part be attributed to using an efficient RoI step, ie the Canny Edge Detector. A similar method to detect corrosion spots will be presented in the next quarter.

References

[1] M. K. Hu, "Visual Pattern Recognition by Moment Invariants," *IRE Trans. Inf. Theory*, 1962.

[2] H. Bilen and A. Vedaldi, "Weakly Supervised Deep Detection Networks," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016.

[3] J. R. R. Uijlings, K. E. A. Van De Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," *Int. J. Comput. Vis.*, 2013.