Quarterly Report No. 1

Slow Crack Growth Evaluation of Vintage Polyethylene Pipes

DOT Project No.: 643
Contract Number: DTPH5615T00007

Reporting Period: 1st Quarterly Report

Date of Report: December 31, 2015

Prepared For:
U. S. Department of Transportation
Pipeline and Hazardous Materials Safety Administration

Prepared By:
Ernest Lever
Principal Investigator
Gas Technology Institute
ernest.lever@gastechnology.org

Project Manager:
Andrew Hammerschmidt
Program Director
andrew.hammerschmidt@gastechnology.org

Subcontractors:
Arizona State University
University of Colorado-Denver

Gas Technology Institute (GTI)
1700 S. Mount Prospect Rd.
Des Plaines, Illinois 60018
www.gastechnology.org

1. Objectives

This collaborative program will provide an integrated set of quantitative tools that will provide a structured approach to reducing operational risk in vintage plastic distribution systems susceptible to Slow Crack Growth failures. A novel endoscopic structured light scanning tool will be developed and prototyped for internal inspection of small diameter plastic pipe. The data generated by the tool will be properly reduced to essential parameters to be synthesized with additional available system information including external conditions, inspection and leak records, historic data, and subject matter expertise into a fitness for service evaluation. This assessment will include a probabilistic estimate of the remaining effective lifetime of individual segments of vintage plastic pipe and a yes/no determination of whether a short-term pressure test is capable of validating the maximum defect size in the system. The Bayesian network methods employed are ideally suited to evaluating interacting threats, investigating root causes, and predicting the effect of mitigation strategies based on conditional probabilities calculated from available data.

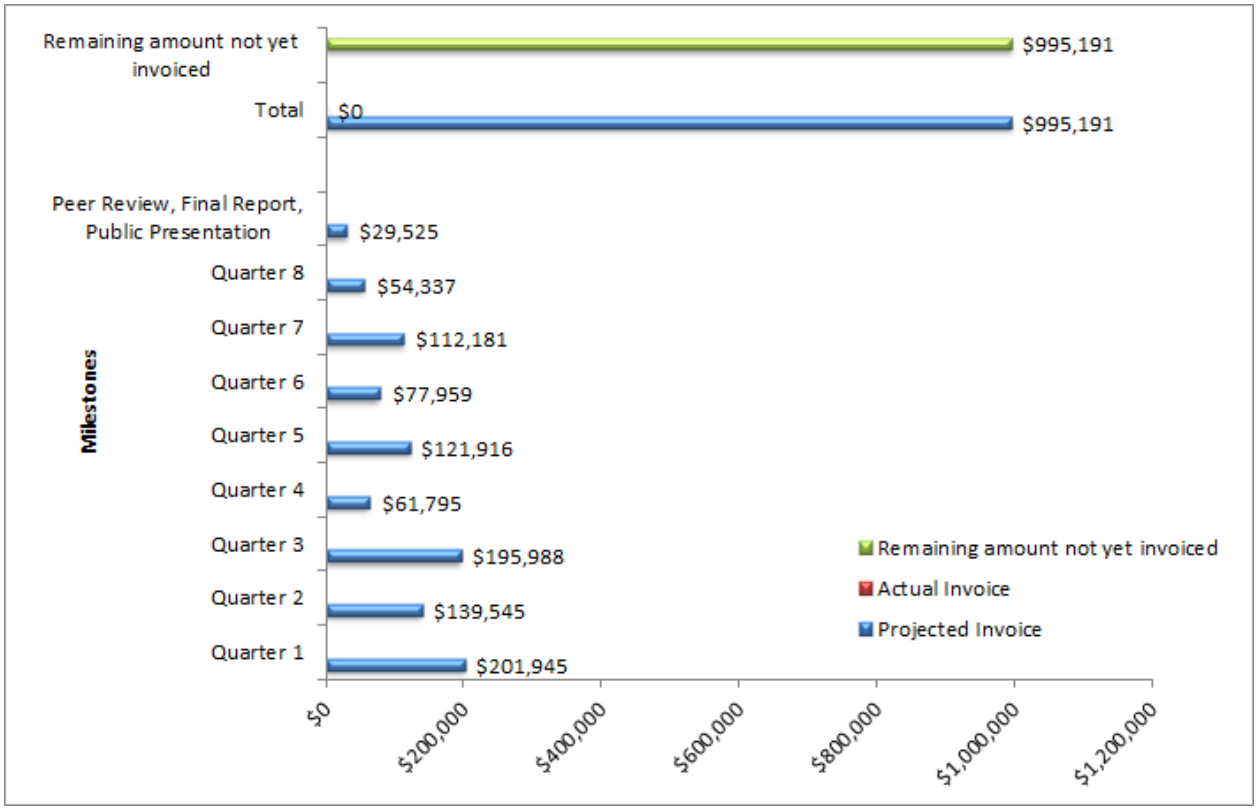
2. Work Completed During Reporting Period

Work in this quarter included:

- Task 1: Probabilistic Decision Support System Design – work was initiated on reviewing existing models, developing the appropriate ontologies, as well as designing the system architecture.
- Task 2: Probabilistic Decision Support System Development - An initial version of Bayesian network was developed. Work also commenced on developing the Enterprise Decision Support System (EDSS) framework.
- Task 3: Structured Light Scanning Method Development – The University of Colorado-Denver initiated progress on the 3D optical inspection system, including experimental work on the single color laser source and endoscopic camera for data collection. 3D image reconstruction of the data was also initiated.
- Task 4: Bayesian Methods Development – Arizona State University performed a review of the current Bayesian network and image processing techniques. A stabilizing device to hold the camera in the pipe was developed and prototyped. Finally, GPU programming was also proposed for calculation and data processing for our scenario.
- Task 6: Project Initiation Activities – The Technical Advisory Committee (TAC) was formed during the first quarter, and the formal Project Kick-off meeting was held on December 4th.

3. Payable Milestones and Planned Activity

Figure 1 shows the quarterly payable milestones to the end of the project.



Plans for Future Activity (Project Quarter #2)

- Task 1: Probabilistic Decision Support System Design – work on reviewing existing models, developing the appropriate ontologies, as well as designing the system architecture will be completed.
- Task 2: Probabilistic Decision Support System Development – work will continue on development of the Bayesian network and the Enterprise Decision Support System (EDSS) framework.
- Task 3: Structured Light Scanning Method Development – The University of Colorado-Denver will continue making progress on the 3D optical inspection system, including additional experimental work on the single color laser source and endoscopic camera for data collection. The literature review will be completed.

- Task 4: Bayesian Methods Development – Arizona State University will ocntinue work on Bayesian network and image processing techniques.
- Task 6: Project Initiation Activities – The Technical Advisory Committee (TAC) will meet during the second quarter.

4. Technical Status

4.1 Summary

GTI, Arizona State University, and University of Colorado-Denver made solid technical progress on Tasks 1 through 4, as well as formed the Technical Advisory Committee in Task 6.

4.2 GTI Task 1&2

4.2.1 Introduction

Many pipeline incidents are the result of multiple, interacting causes, not a single threat. Individual threats can each be at "acceptable" levels but when overlaid result in a significant threat to the pipeline or even a failure. However, there is a lack of comprehensive research in addressing such risks in a complex environment. In this task, we will develop an enterprise decision support system platform by employing Bayesian network and semantic ontology to model the interactions between factors leading to threats or failures. Ontology describes domain knowledge and is designed by subject matter expert. Bayesian network allows calculation of the probability of failures and classification of them into various categories based on risk levels. It is ideally suited to evaluating interacting threats, investigating root causes, and predicting the effect of mitigation strategies based on conditional probabilities calculated from available data. This assessment will include a probabilistic estimate of the remaining effective lifetime of individual segments of vintage plastic pipe and a yes/no determination of whether a short-term pressure test is capable of validating the maximum defect size in the system. The progress in the 1st quarter is presented in the following sections.

4.2.2 Ontology

An ontology is a formal description of some application domain. It defines the concepts and entities involved, as well as the relationships between them. Precisely, it is a formal specification of a shared conceptualization. It is founded on Description Logics and expressed using languages defined by the Web Ontology Language (OWL) language.

In this work, we have represented subject matter expertise of pipeline threats in ontological format. Figure 4.2.1 illustrates an example of such an ontology describing interactions between factors contributing to threats in vintage pipelines. Some of the factors are residual stress, wrinkle bends/miter bend, fabrication/weld quality, 3rd party damage, external corrosion, internal corrosion, stress corrosion cracking, soil and other superimposed stresses, and unknown, hidden, incorrect operations threats.

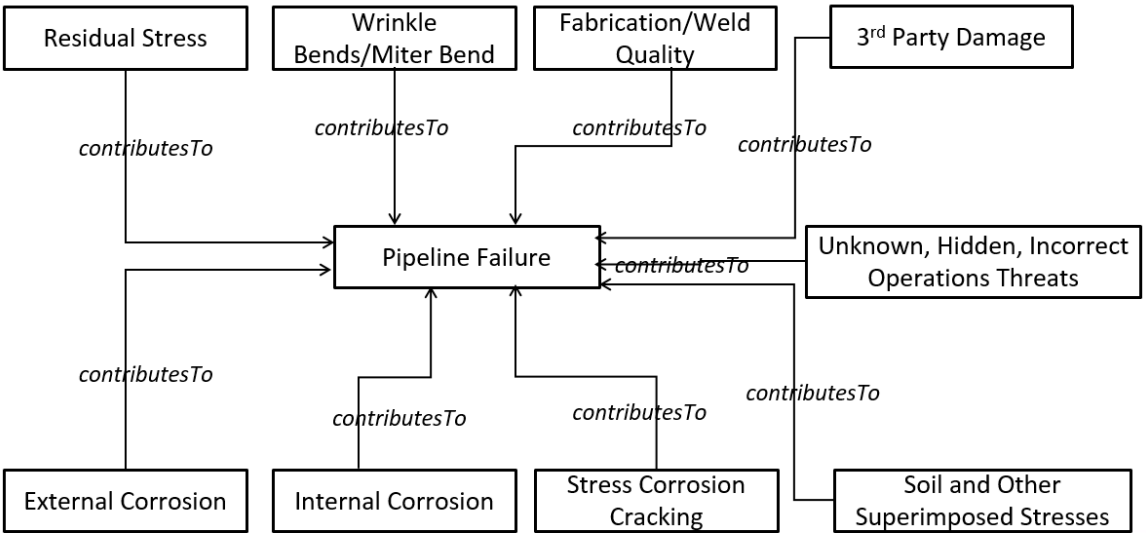


Figure 4.2.1: Example of ontology representing threats leading to pipeline failure

We followed an ontology design standard called *Event-Model-F* [30] which gives a formal model of events. Event describes an action at a certain time and location. The model is based on the foundational principles of ontology and provides comprehensive support to represent time and space, objects and persons, as well as causal and correlative relationships between events. Figure 4.2.2 is an ontology of failure events described by using a combination of causality and observation patterns. It represents cracked gas pipeline event leading to the gas leakage event which is reported by an operator.

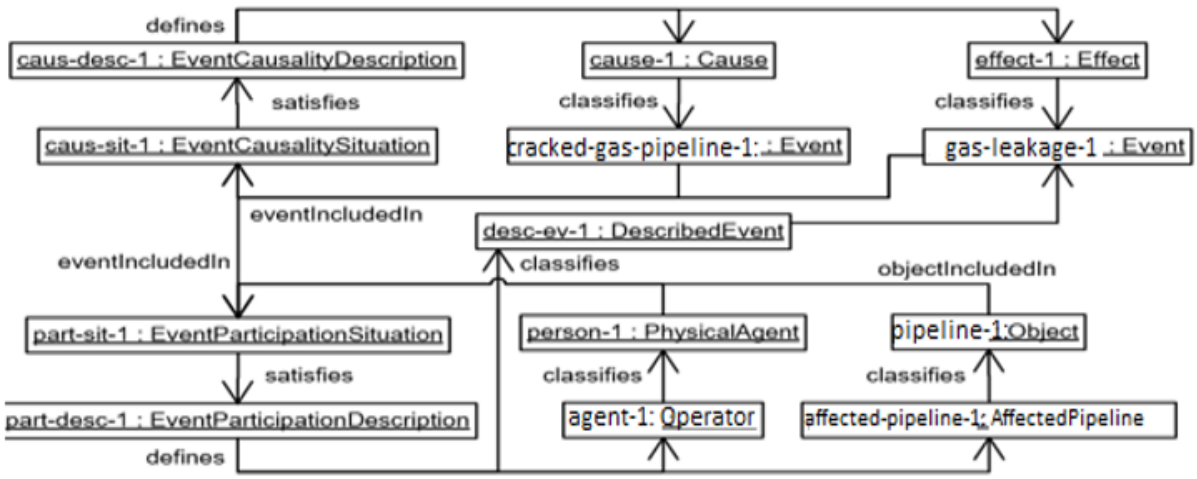


Figure 4.2.2: Example ontology illustrating causality in gas pipeline distribution

4.2.3 Bayesian Network

A Bayesian network is a graphical probabilistic model in which nodes represent variables of interest and edges represent the dependency between variables. It enables calculation of threat interaction levels and severity of risks in vintage gas pipelines. As a part of Bayesian network model development, we will build causal network, subject matter expert network, historical data network, and inspection reports network.

In this quarter, an initial version of Bayesian network was developed as follows. The first step was for the subject matter experts (SMEs) to identify factors that impact the performance of vintage plastic gas distribution pipelines. 32 factors are identified, as shown in Figure 4.2.3, as the nodes of the Bayesian network.

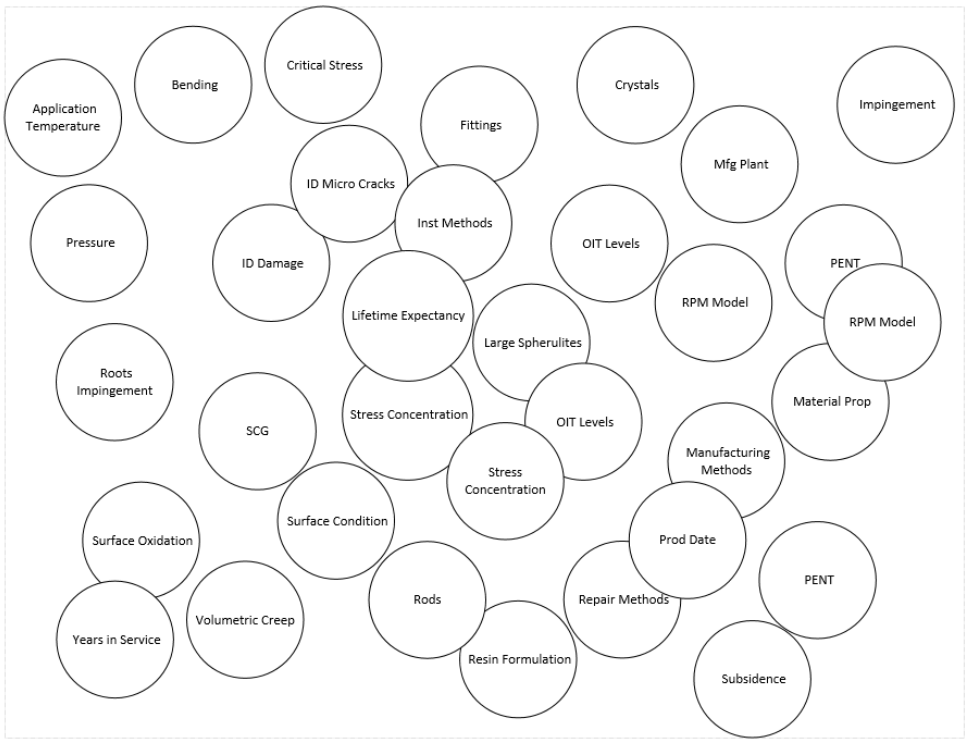


Figure 4.2.3: Factors selected by SMEs

The second step was to draw edges between the nodes. Out of a theoretical possible combination of 4,294,967,296 interactions between the 32 factors, SMEs then selected 65 plausible interactions to construct the edges of the initial Bayesian network as shown in Figure 4.2.4. Two simple graphical analyses were then run. First, betweenness centrality, which is equal to the number of shortest paths from all vertices to all others that pass through that node, was calculated for every node. It is shown in Figure 4.2.5. Second, degree centrality, which is the number of ties a node has to other nodes, was calculated for each node as shown in Figure 4.2.6. Larger red dot denotes higher centrality value in both Figures 4.2.5 and 4.2.6. The product of degree centrality and betweenness centrality gives a composite ranking, which then helps to identify the most important factors. Figure 4.2.7a and 4.2.7b shows the composite ranking of nodes in the Bayesian network. As expected, the top five factors with the highest composite rankings in decreasing order are Stress concentration, Manufacturing methods, Bending, RPM model, and Surface condition.

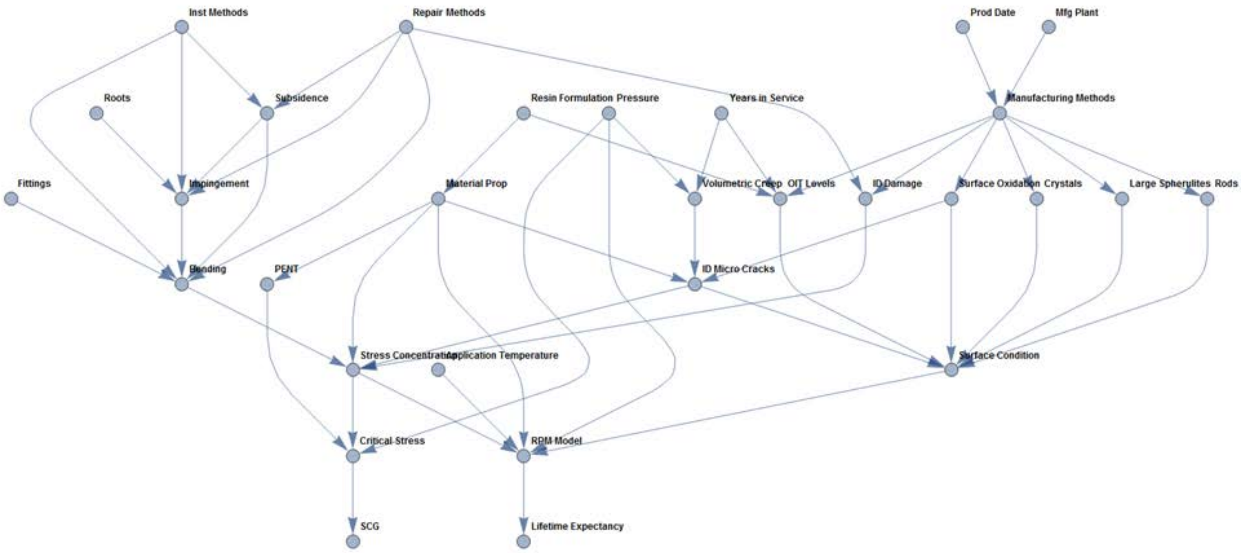


Figure 4.2.4: Initial Bayesian network graph selected by SMEs

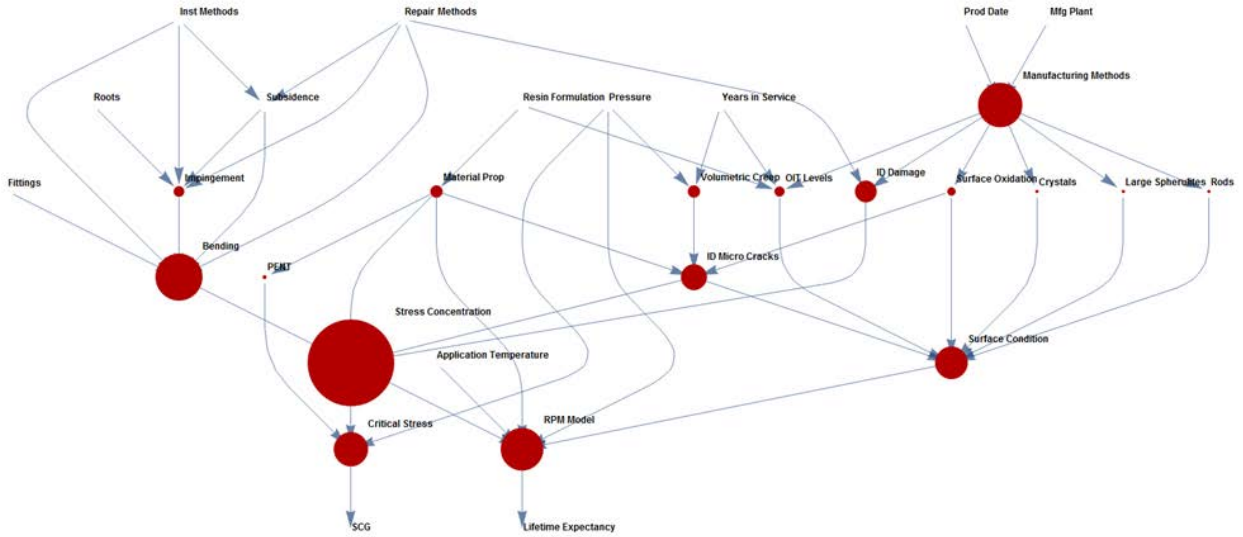


Figure 4.2.5: Betweenness centrality of the Bayesian network graph

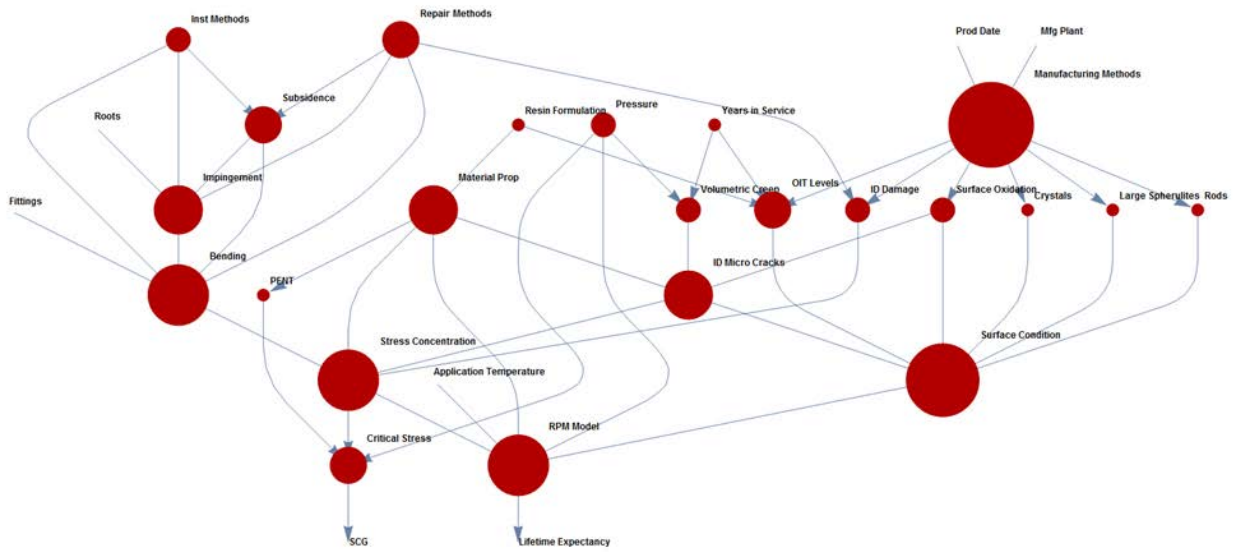


Figure 4.2.6: Degree centrality of the Bayesian network graph

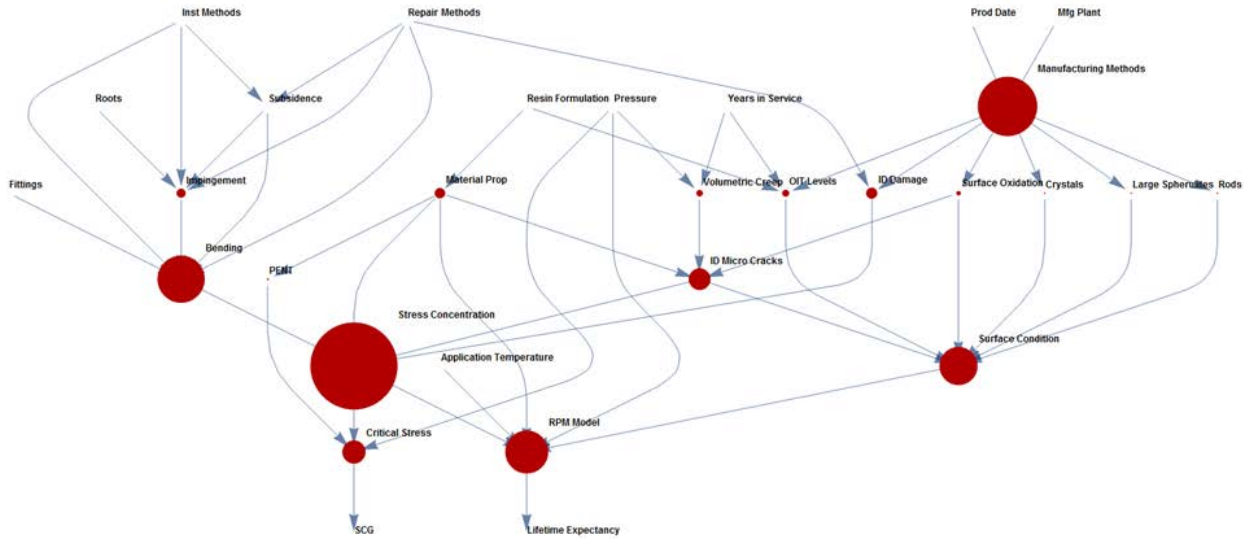


Figure 4.2.7a: Composite Ranking of the Bayesian network graph

Stress Concentration	51.
Bending	27.5
Manufacturing Methods	26.
RPM Model	25.
Critical Stress	20.
Surface Condition	19.
ID Micro Cracks	15.
ID Damage	12.5
Volumetric Creep	7.
Material Prop	7.
Impingement	6.
OIT Levels	5.6
Surface Oxidation	4.6
PENT	2.
Rods	1.6
Large Spherulites	1.6
Crystals	1.6
Years in Service	0.
Subsidence	0.
SCG	0.
Roots	0.
Resin Formulation	0.
Repair Methods	0.
Prod Date	0.
Pressure	0.
Mfg Plant	0.
Lifetime Expectancy	0.
Inst Methods	0.
Fittings	0.
Application Temperature	0.

Betweenness Centrality

Manufacturing Methods	8
Surface Condition	7
Stress Concentration	6
RPM Model	6
Bending	6
Material Prop	5
Impingement	5
ID Micro Cracks	5
Subsidence	4
Repair Methods	4
OIT Levels	4
Critical Stress	4
Volumetric Creep	3
Surface Oxidation	3
Pressure	3
Inst Methods	3
ID Damage	3
Years in Service	2
Rods	2
Resin Formulation	2
PENT	2
Large Spherulites	2
Crystals	2
SCG	1
Roots	1
Prod Date	1
Mfg Plant	1
Lifetime Expectancy	1
Fittings	1
Application Temperature	1

Degree Centrality

Stress Concentration	306.
Manufacturing Methods	208.
Bending	165.
RPM Model	150.
Surface Condition	133.
Critical Stress	80.
ID Micro Cracks	75.
ID Damage	37.5
Material Prop	35.
Impingement	30.
OIT Levels	22.4
Volumetric Creep	21.
Surface Oxidation	13.8
PENT	4.
Rods	3.2
Large Spherulites	3.2
Crystals	3.2
Years in Service	0.
Subsidence	0.
SCG	0.
Roots	0.
Resin Formulation	0.
Repair Methods	0.
Prod Date	0.
Pressure	0.
Mfg Plant	0.
Lifetime Expectancy	0.
Inst Methods	0.
Fittings	0.
Application Temperature	0.

Composite Ranking = Betweenness Centrality * Degree Centrality

Figure 4.2.7b: Composite Ranking of the nodes in Bayesian network graph

The third step is to construct conditional probability tables for the factors in the Bayesian network. Three methods were identified for the purpose: FEM analysis, Data collection and analysis, and Experimentation and observation as shown in Figure 4.2.8. We used an extensive historic hydrostatic test data covering all vintage pipelines to construct the preliminary conditional probability table. Figure 4.2.9 demonstrates this ongoing work. The initial study was conducted on a set of 105 slow crack growth failures in pipe that were not associated with fittings or other known stress risers. The pipes were all extracted from a gas distribution system in 2010 -2011. They had been in service since 1972, 1973 and 1974. A combination of simpler statistical models, subject matter expertise and historical observations was used to estimate initial graphical structure and probabilistic distribution of a Bayesian network model. This process is represented in ontology illustrated by Figure 4.2.10. Statistical models help to understand various levels of threat interactions between factors of interest that may lead to failure. Some of the modeling methods, evaluated in this preliminary work, are rate process method, fault tree analysis, event tree analysis, surface feature model, finite element method, risk score model, etc.

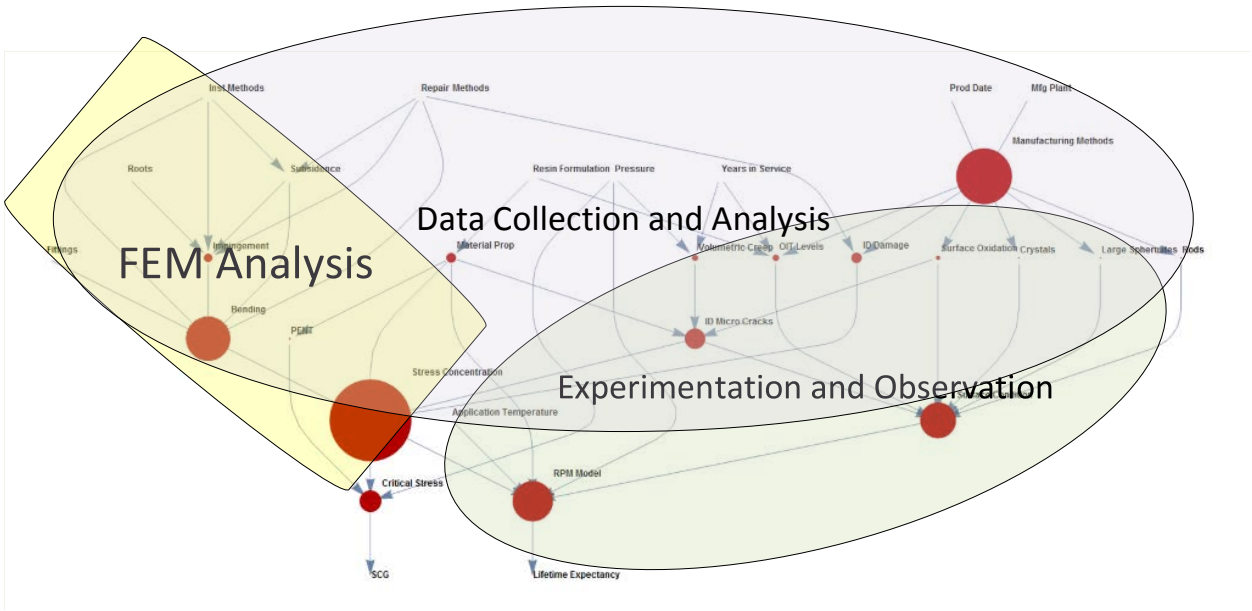


Figure: 4.2.8: Methods to derive the conditional probablity tables in the Bayesian network

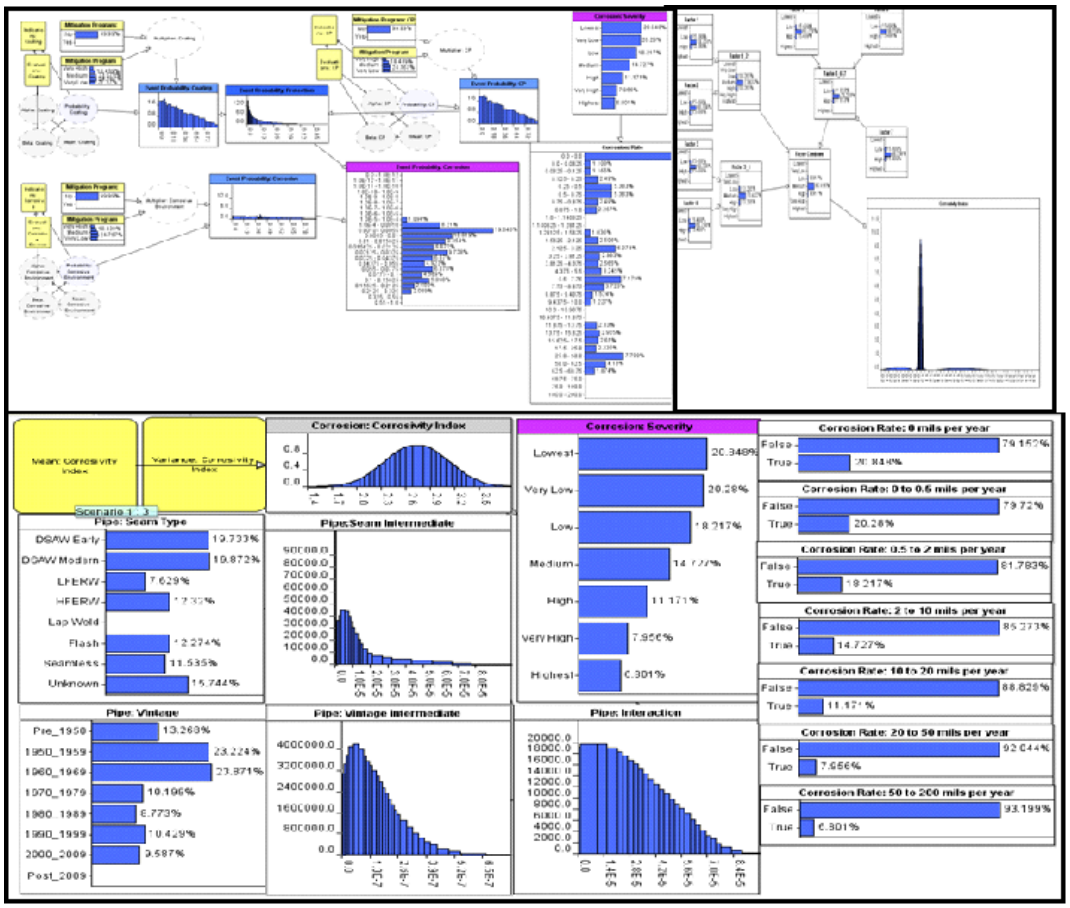


Figure 4.2.9: Preliminary conditional probability distributions

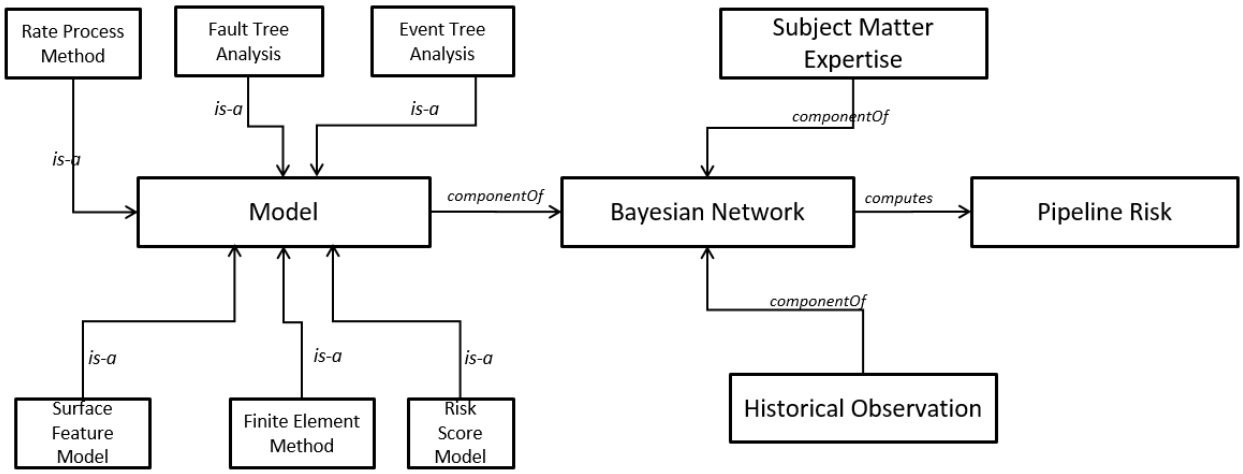


Figure 4.2.10: Ontology describing the models, subject matter expert's knowledge, historical observation, and Bayesian network

4.2.4 Enterprise Decision Support System

Enterprise Decision Support system (EDSS) enables operators, regulators and utilities to apply science and engineering methods on heterogeneous data sources related to pipeline distribution for fitness of service evaluation. The data sources may contain available system information including external conditions, inspection and leak records, historic data, customer data, and subject matter expertise. EDSS has models that calculate threat interaction levels and their severity. It provides a method to continuously monitor threat interactions and flag concerns at trigger points.

The conceptual design of EDSS is given below.

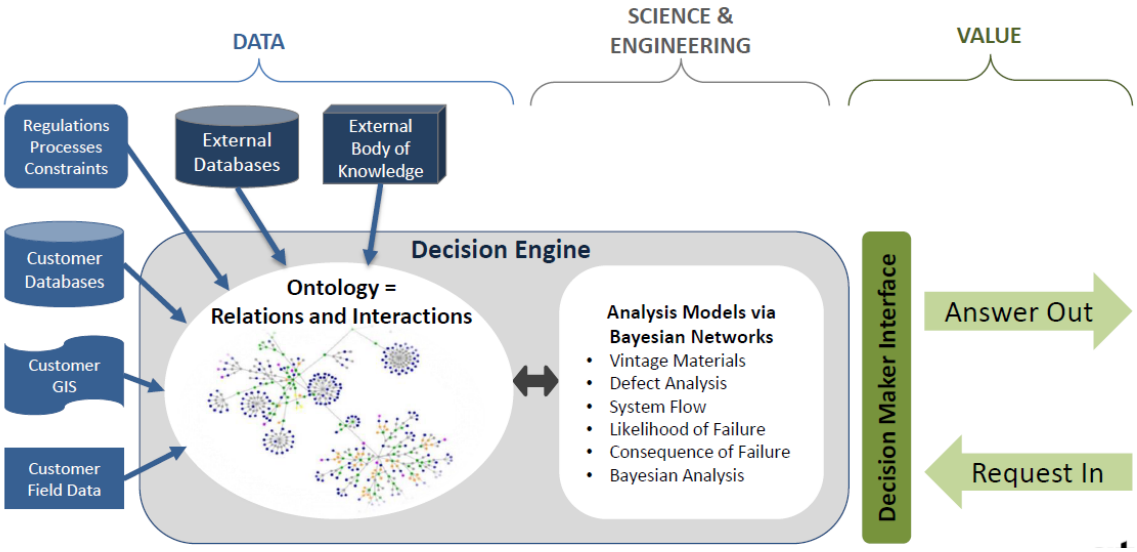


Figure 4.2.11: Conceptual design of Enterprise Decision Support System

The UML use case diagram of the EDSS is given in Figure 4.2.12. Operators, Regulators, Utilities and Subject matter experts are the stakeholders in EDSS.

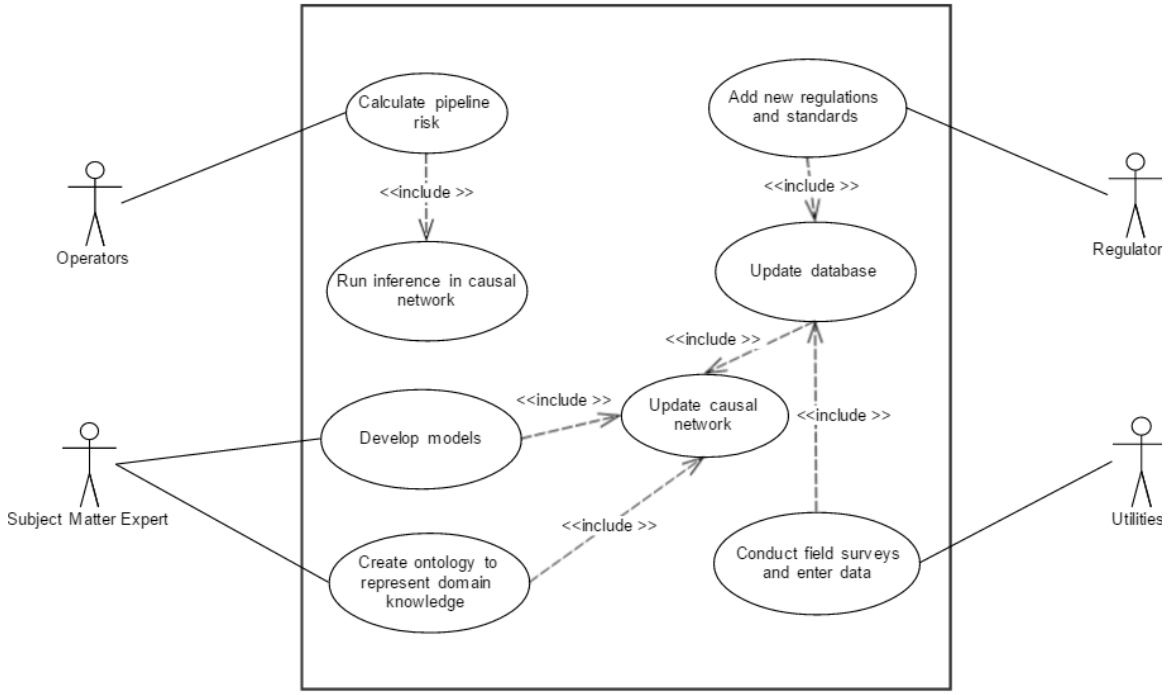


Figure 4.2.12: Use case diagram of EDSS

We now present the Software-as-a-Service architecture of the EDSS platform in Figure 4.2.13. It is a real-time, cloud-based scalable system hosted in Microsoft Azure. The heterogeneous dispersed data are stored in its big data source and semantically enabled models are deployed into the EDSS platform. EDSS is being built with open-source technologies mostly licensed under Apache 2.0 license. Two open standards have been identified for data communication in and out of EDSS: resource description framework (RDF) and JSON-LD (JavaScript Object Notation for Linked Data).

- RDF is a language for representing semantic information about objects in the World Wide Web. It is a graph based data model which provides grammar for its syntax and supports query against the model. Several design tools, including Protégé, are available for designing RDF ontology. Due to the availability of design tools and ease of usage, we have selected RDF as an ontological standard to represent subject matter experts' knowledge.
- JSON-LD is a specialized RDF syntax. As JSON-LD is also JSON document that represents an instance of an RDF data model, it fits well with the schema- less NoSQL database. Therefore, we have selected JSON-LD as communication standard between applications or models.

ISO 19465/AMQP 1.0 has been identified as an appropriate open standard messaging protocol for EDSS to efficiently transfer information within and between utilities or government agencies. It enables cross-platform applications to be built using brokers, libraries and frameworks from different vendors. AMQP 1.0 is supported in Microsoft Windows Azure and Redhat Linux operating systems. Some of its early backers are US Department of Homeland Security, Microsoft Corporation, VMware Inc, Cisco Systems, Mitre Corporation, Bank of America, Goldman Sachs, etc.

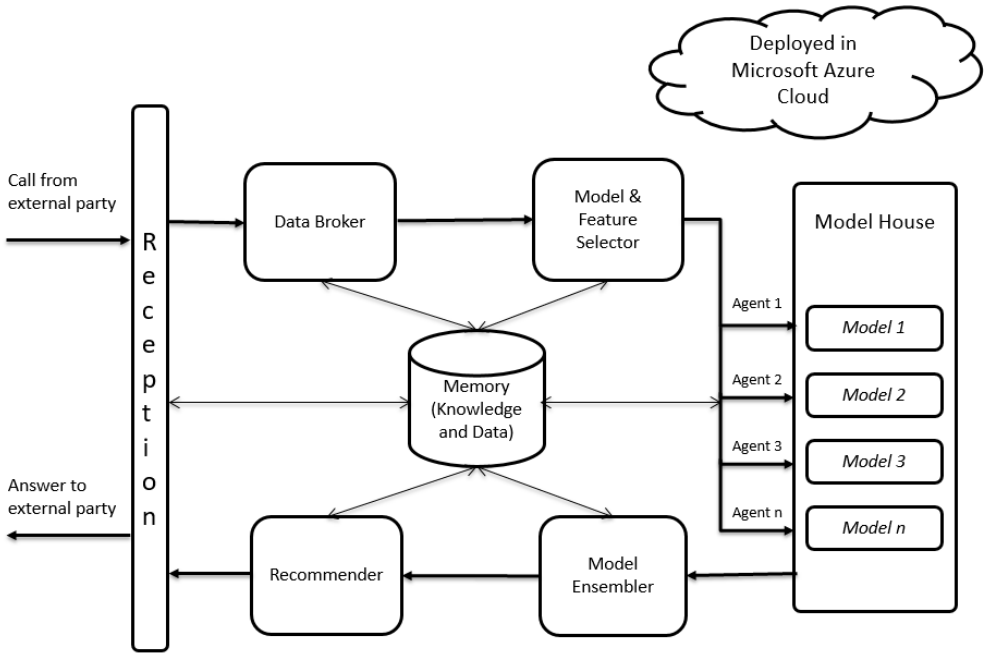


Figure 4.2.13: Architecture of Enterprise Decision Support System

There are nine components of EDSS platform.

Reception: It is ISO 19464/AMQP 1.0 standard communication bus which interacts with external services. It accepts queries and data as input and returns the result of causal Bayesian inference.

Data broker: It fetches information from the database. An ontological method will facilitate such data fetch.

Database/Memory: Big database technologies such as NoSQL, Graph database and Apache Hadoop are used as the database technology. NoSQL and graph database overcome the limitation of structured relational database systems by allowing schema-less data. Consequently, a variety of data in any form such as customer GIS, customer database, regulatory requirements, standards, constraints, events, sensor logs, etc are supported within the same data store. Apache Hadoop will be used as a data warehouse to ultimately store the data which may be used for analytics in future.

Model and feature selector: It performs two tasks. First, relevant single or multiple models are selected based on the query and data. Second, it identifies the important features and selects them from data by performing statistical analysis.

Model: A model is a description of the system using mathematical concepts and language. Examples of models are Bayesian network, neural network, fault tree analysis, event tree analysis, surface feature model and finite element method, etc. Models developed with any of the programming languages such as JAVA, R, MATLAB and Python are supported.

Model House: It stores a collection of models. Graphical processing unit and central processing unit provides scalable and distributed cloud computing power to the models as required, thus supporting simple mathematical model to compute intensive image processing models.

Agent: An agent is an autonomous entity which observes events in an environment and directs its activity towards achieving goals. Each model has its own dedicated agent. The agents execute models asynchronously in parallel. In addition, it prepares data in specific format for the selected model and pulls parameter values for models from memory.

Model Ensembler: It uses multiple learning algorithms to obtain better result than could be obtained from any of the constituent models. It decides the best solution from the results of multiple models.

Recommender: It formats the final result in an appropriate format and logs that information in memory.

The design of EDSS platform as a Software-as-a-Service (SaaS) enables it to communicate with other enterprise or government applications, thereby increasing its usability and a value addition to utilities, operators, regulators and public. Figures 4.2.14 and 4.2.15 illustrate this usage. In the next quarter, we plan to develop toolsets to help design gas distribution standards and to enable regulators to add regulatory requirements in EDSS platform as shown in Figure 4.2.16.

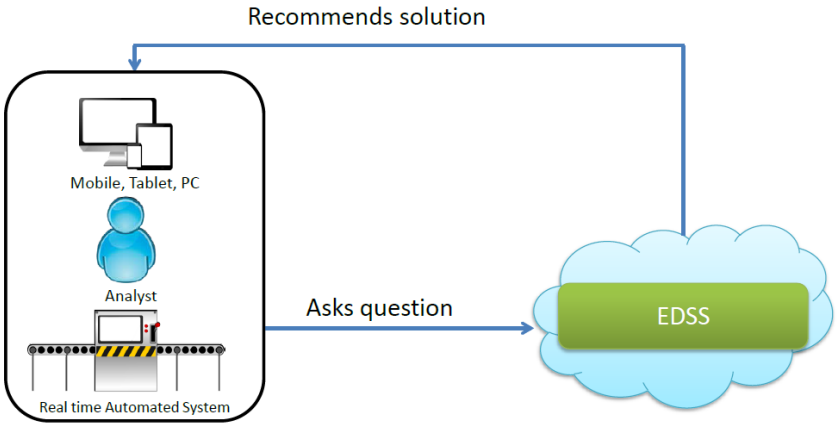


Figure 4.2.14: Use case of EDSS platform

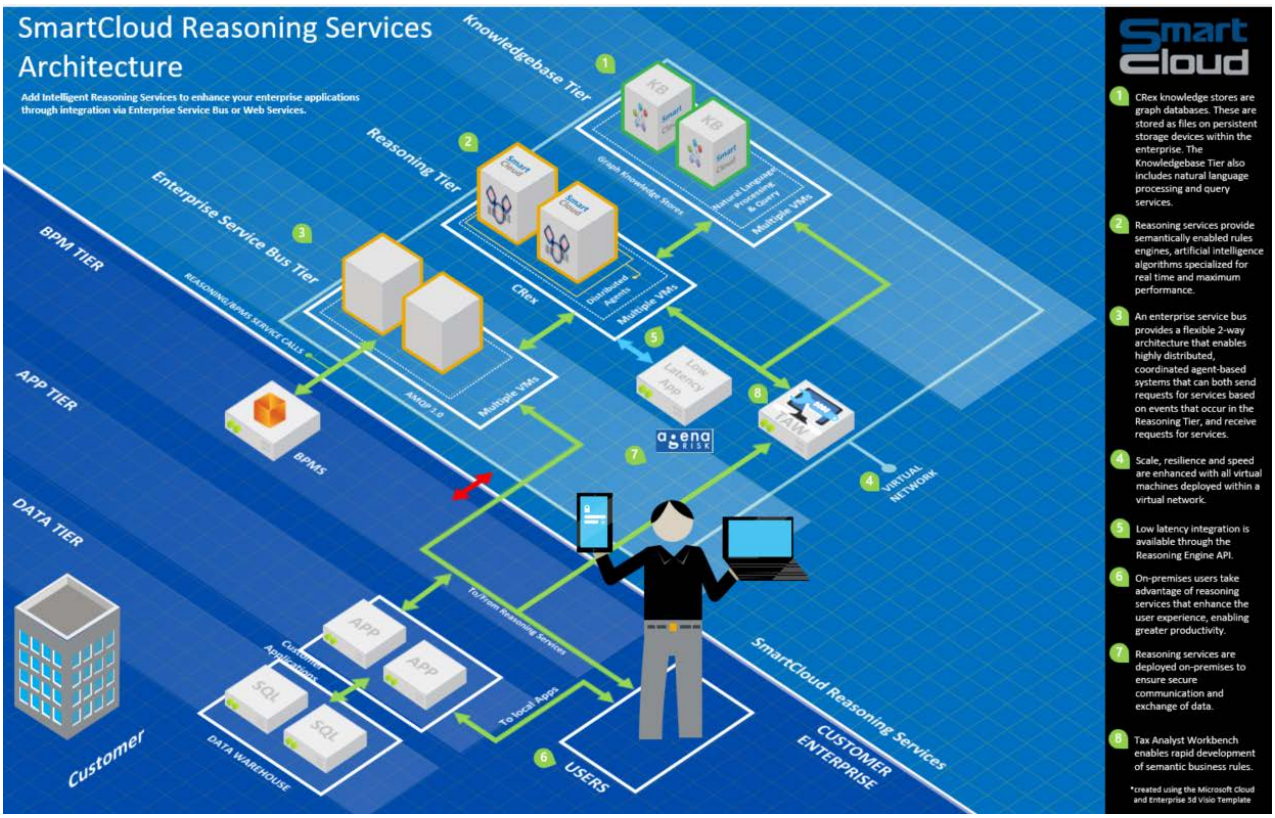


Figure 4.2.15: Partnership with 3rd parties to integrate EDSS platform with their services

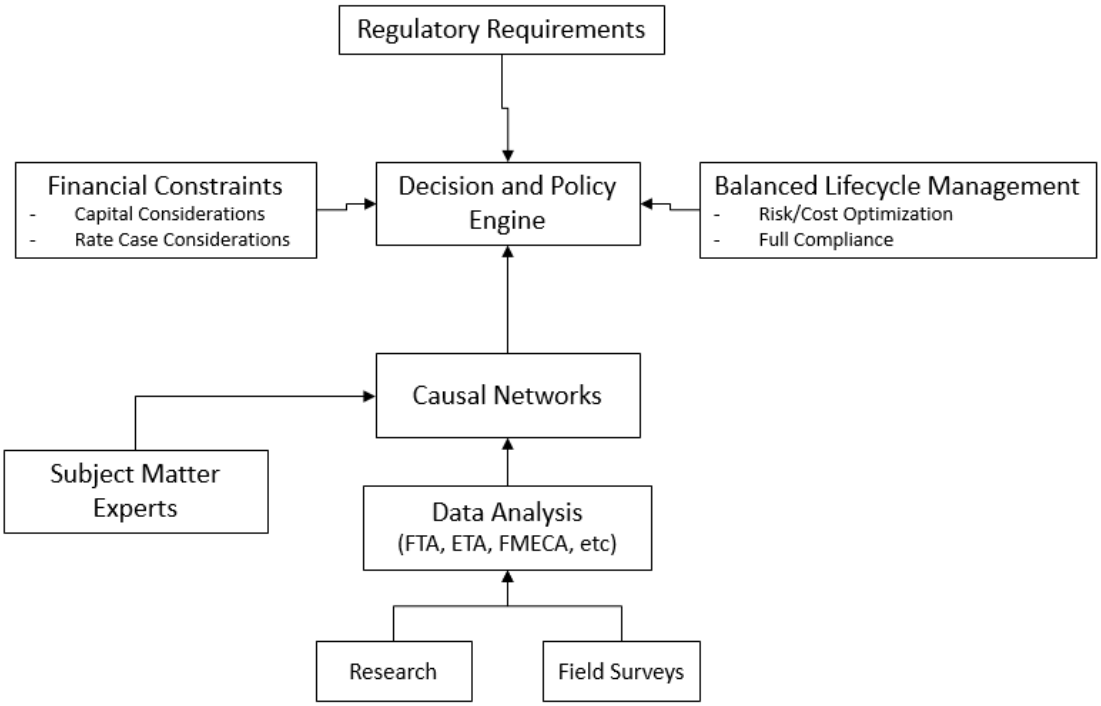


Figure 4.2.16: Regulations and Standards guiding decision making process in EDSS

4.2.5 Data Entry

In this quarter, we designed smart form for data entry purpose. Smart form is an electronic form with capability beyond a traditional data entry interface with a fixed schema. The fields in the form are generated dynamically based on the content and context of the data entered so far. Subject matter expertise, represented as ontology, guides the process of data entry. It will guide the operator through the relevant data gathering stages and ensure only valid and relevant data is fed directly into the appropriate models. It is designed to overcome human data entry errors, which are the most common source of bad data, and to collect as much complete data as possible.

4.2.6 Simulation

EDSS platform provides a set of toolsets for simulation and modeling. It has access to heterogeneous big data sources including simulated test data, manufacturers' data, field failure data, research reports, academic literature, standards and regulations, and subject matter expertise. Operators or analysts may run simulation with models that include application specific handbook models, probabilistic risk models, and calculators. Multiple competing models may be deployed and run in parallel to evaluate the results.

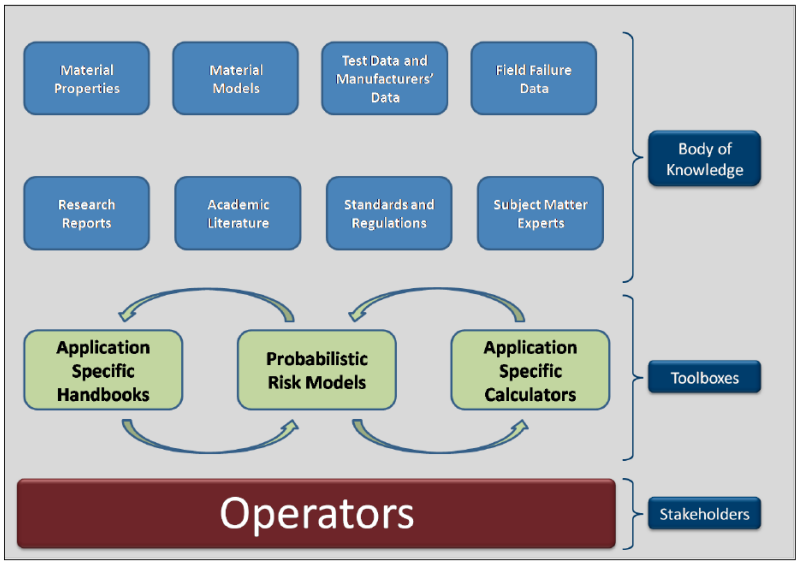


Figure 4.2.17: Simulation model toolboxes with access to heterogeneous data sources

4.2.7 Insights Interface

Insights interface is designed as a web-based dashboard that operators or regulators can configure to gain key insights into the information they want including vital details about the data, regulations and standards. Furthermore, it allows users to select multiple risk models,

enter parameters, and run them in parallel to display the comparative results as graphs and visuals.

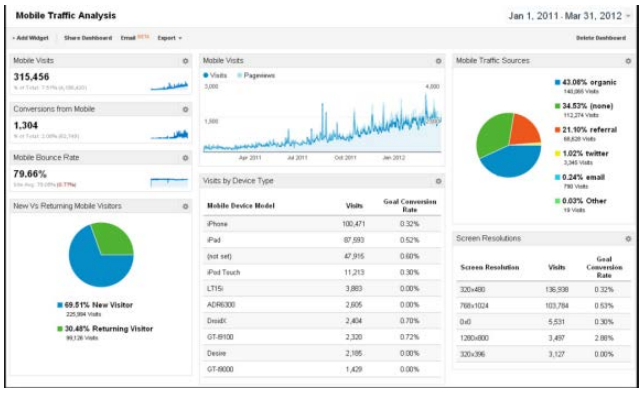


Figure 4.2.18: A mockup version of insights interface dashboard

4.3 CU-Denver Task Work Summary

4.3.1 Introduction

Identification and classification of current vintage pipeline inner wall damage precursor are of critical importance. However, there are limited success in sensing and characterizing the long lengths of deteriorating plastic pipelines with high probability of detection (POD), and the capability of the currently available accelerometers and imaging technologies that can be miniaturized and integrated into smaller CTS size pipes for a fast scan is questionable and needs a systematic assessment. There are several known inner pipe wall surface conditions that increase the probability of premature failure due to slow crack growth. The major drivers of premature failure due to slow crack growth are bending stresses due to tight bend radii, impingement and fittings. Damage is also introduced by pipe squeeze-off during maintenance operations. These conditions identification has been investigated by various nondestructive evaluation (NDE) techniques, such as direct visual/optical methods using CCD cameras, ultrasonic testing, liquid-coupled acoustic measurement (e.g. sonar) and laser based surface inspection approaches including light detection and ranging (LiDAR) and laser topography [1-3]. However, these currently available technologies suffer either from low-sensitivity and resolution for small damage precursors, or complex settings and large system footprint that makes it incapable for plastic gas distribution pipes with much smaller diameters. In this task 4.3, led by the PI at CU-Denver, we will develop a multi-spectral and miniaturized optical sensing platform for imaging the smaller CTS size pipe inner walls with high sensitivity and specificity of damage identification. The literature review and progress in the 1st quarter are presented in the following sections.

4.3.2 Principle of 3D optical inspection system

3D profiling has been widely used and has a lot of applications in manufacturing and medical field. The principle of capturing 3D image is basically the triangulation mathematically. This technique depends on two basic and critical components in the system, which are camera and projector. The optical projector/source sends structure light to the object, where a typical structure light pattern is shown in Figure 4.3.1. The benefit of using this light is to computationally capture the depth information from the 3D object from a relatively simple light projection. In the literature, a pinhole model has been used to capture the light that reflected from the object.

$$\tilde{m}^c \propto K^c [R^c \; t^c] \tilde{M}^w \qquad \text{(eq. 4.3.1)}$$

where $\propto$ stands for the equality up to scale factor, $[R^c \; t^c]$ is extrinsic parameters matrix and K^c is the camera parameters whose matrix form is shown in eqn. (4.2.2):

$$K^c = \begin{bmatrix} \propto & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \text{(eq. 4.3.2)}$$

$u_0 \; and \; v_0$ are coordinates of the principle point, while $\propto \; and \; \beta$ are the focal lengths.

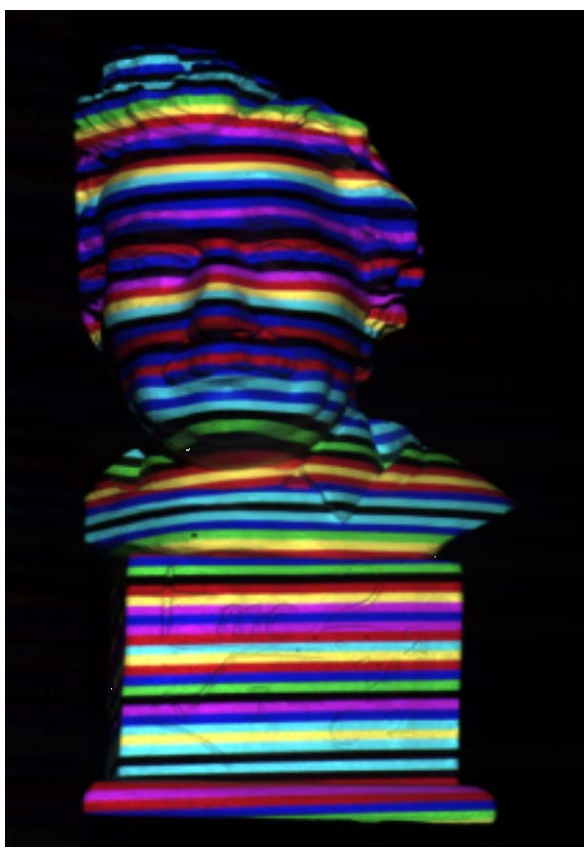


Figure 4.3.1: Example of a structured light

By using the structured light illumination, the collected image by the camera will be reconstructed into 3D data. Figure 4.3.2 illustrates the triangulation system by camera and light source. In this task, by taking the advantage of the endoscope technique, the innovative imaging principle of using the multi-color multi-ring is considered to increase the resolution and sensing speed for possible in-line inspection (ILI) integration. Although there are many kinds of structured light methods, Single Shot (SS) technique is one of the most used techniques and we will adopt it for the first prototype development. Since the signal shot just needs one image of the scene to reconstruct 3D data, it will simplify the sensor hardware design as well as reduce the image processing load. In other literature in medical field, one design of structured light endoscope needs camera and a slide projector to collect the 3D data, which will also be investigated in this task. Figure 4.3.3 shows the endoscope technique using for clinical applications at the diameter of 3.8mm.

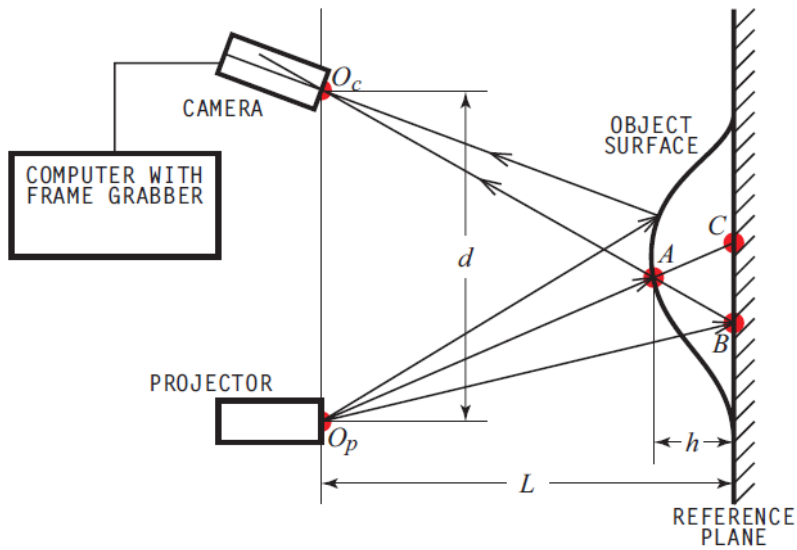


Figure 4.3.2: Illustration of a 3D structured light measurement

4.3.3 Experimental work

By taking advantage of the triangulation, in this R&D phase, the single color laser source and endoscopic camera have been used to collect the optical data in the laboratory testing environment. For prototype I, the laser source was projected to the pipe inner wall as a red ring that is shown in Figure 4.3.4(b) and 4.3.4(c). Good contrast was achieved with damage induced distortion can be easily extracted as shown in the figure. However, because of the use of small view angle camera, the output image was not a full circle, which will be improved in the 2nd quarter.

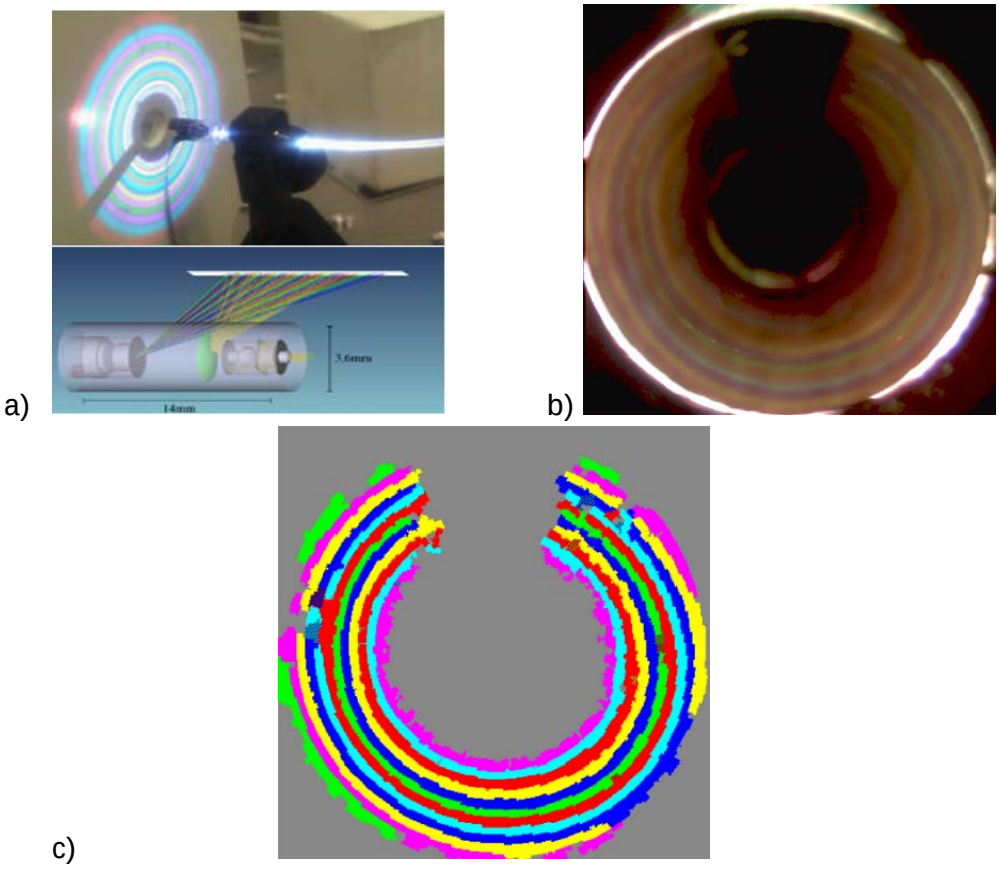
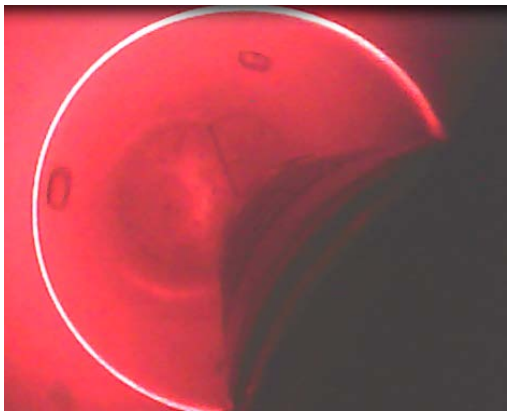


Figure 4.3.3: Example of acquired images with structured light scanning [5] a) a biomedical endoscopic 3D scanner prototype b) raw image; c) decoded results revealing the rich inner wall surface profiling details due to distortions from the ideal multi-color rings in a).

The camera has been installed and integrated with the laser source in same plastic holder design and 3D printed at the CU PI's research group. Further improvement of the design is anticipated in the Q2. ASU will also contribute in the scanning stabilizer design to further reduce the imaging noise. This prototype I has been tested in the pipes with 3 inches diameter whose inner walls have artificially introduced damage with various patterns, shown in Figure 4.3.4 a). GTI is arranging the shipment of pipes with more realistic damage to CU for further testing in Q2. The limitation due to the "non-complete-circle" pattern will be studied and removed by the new design of prototype II in the 2nd quarter.

a)

b)



c)

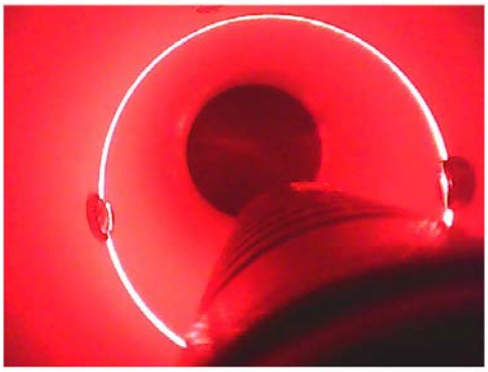


Figure 4.3.4: a) Pipe with artificially introduced defects, b) one frame from the collected data without defects c) one frame from the collected data with two defects

4.3.4. Fast 3D image reconstruction

As mentioned before, the reconstruction process depends on the triangulation between the camera and the projector. In order to determine the shifted location of the projected lines, some intermediate image processing steps are implemented. The complexity of the required steps depends on the type of the used structured light methodology and the amount of noise and artifacts that corrupt the camera image. For this project two types of implementations are

considered. The first one is the single laser ring and the other is the multi-ring multi-color method which will be implemented in this Task.

For the single laser ring, the reconstruction algorithm is applied to either one of the RGB layers that has the maximum intensity for the laser ring or the gray scale version of the image. The extraction process starts with applying a weighted threshold to the image to reduce the noise. After that a two dimensional Sobel filter is applied to extract the edges from the image. To enhance the laser line detection a multiple morphological operations are applied to fill and connect the discontinuities of the laser ring [8]. The laser ring is then extracted and the deformation depth is calculated. Figure 4.3.5 shows a brief description of the implemented algorithm.

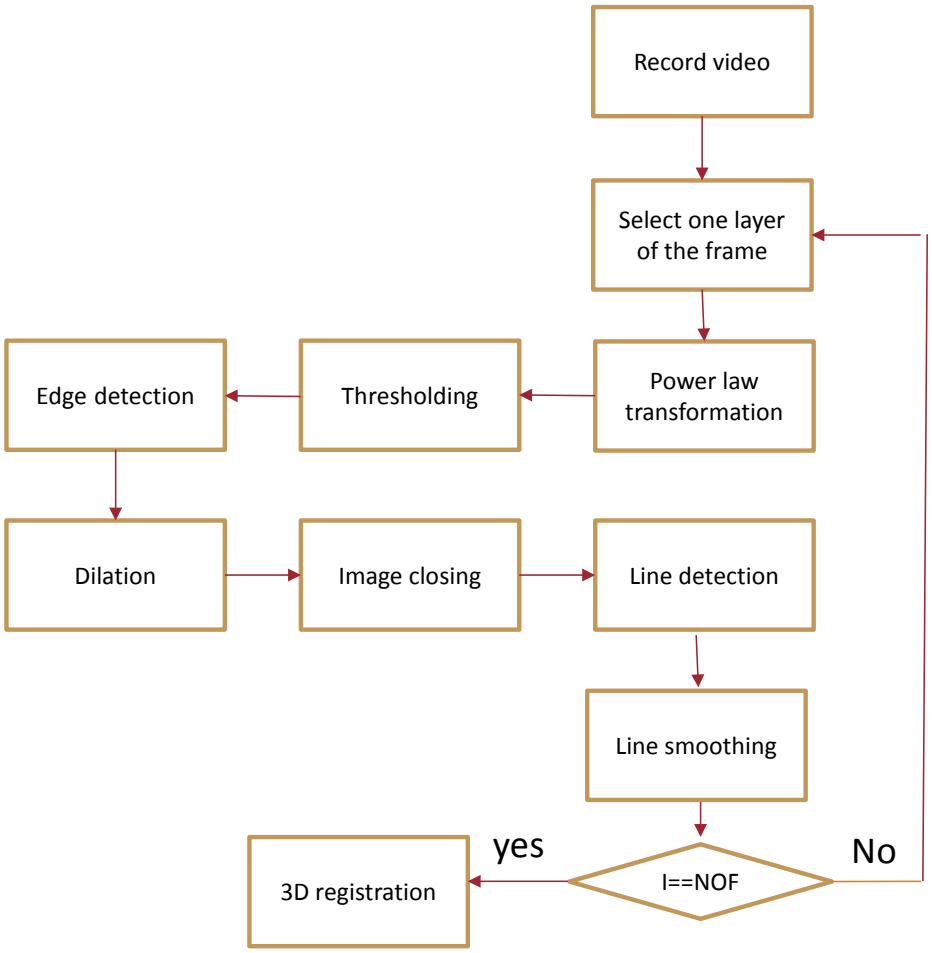


Figure 4.2.5: Reconstruction algorithm for single laser ring

The Multi rings patterns are proposed to increase the resolution and speed of the 3D acquisition. With the introduction of multiple ring in the image, the problem of the ring correspondence arise. One of the proposed solutions to solve this problem is the use of colored

rings. The colors of the rings are coded with a special code to create sets of unique sequences that can be identified correctly in any location inside the image. Many methods are proposed to decode those patterns. One of the methods is to use graph based reconstruction to decode the pattern [5][9]. In this method the image is segmented with watershed transformation. Then a graph diagram is created for the segmented image and the edges are extracted according to this graph diagram.

For both of the described methods, after processing each frame, the depth information is collected and registered to form the final 3D shape. One of the registration methods is performed by using the location of the camera as a reference for the registration process. The other method is to use special image registration algorithm to estimate the location and register the frame. Multiple registration algorithms are proposed to solve this problem, one of the popular algorithms depends on Iterated Closest Point [9]. The problem with those algorithms is that those algorithms fail to deal with flat surfaces due to the lack of any distinctive features [5]. For this reason, a combination of both position data and registration algorithms need to be applied in order to get accurate results.

Preliminary results of the 3D reconstruction for the single laser ring can be seen in Figure 4.3.6. The results show that about 75 percent of the pipe's surface is reconstructed correctly, and the defect can been seen as a bumps or holes on the pipe's walls. The surface reconstruction can be enhanced by using a more advanced setup for the camera and the projector to reduce the amount of the shadows. The results also show some defects due the instability of the camera holder. A more stable holder and a correction algorithm are required to reduce those defects. The projector and camera setup and the required algorithms for the multi ring methodology are still under development and the results will be reported later in Q2 report.

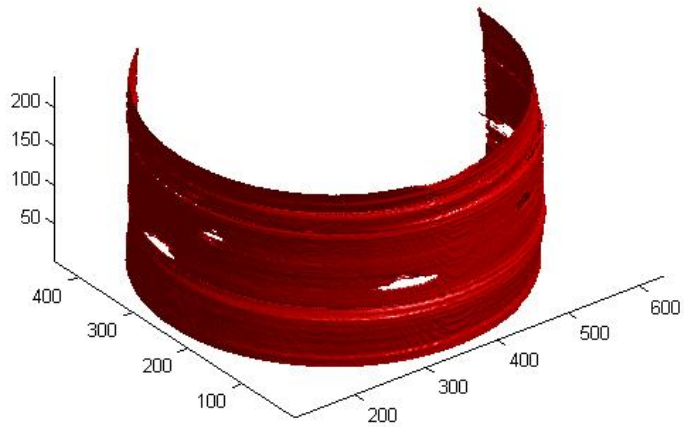


Figure 4.3.6: 3D reconstruction of the internal surface of the pipe

4.4 ASU Task Work Summary

4.4.1 Introduction

Damage diagnosis and remaining life prediction of pipeline infrastructure systems is still a challenging problem despite tremendous progress made during the past several decades, such as the damage accumulation in plastic gas distribution pipes. The goal of our component of the project is to implement a Bayesian network for classification of images taken from inside the pipe and develop a maintenance framework for plastic pipeline systems.

In this report, some review of the current Bayesian network and image processing technique are introduced. A stabilizing device to hold the camera in the pipe is prototyped. GPU programming is also proposed for calculation and data processing in our case.

4.4.2 Bayesian/Maximum Entropy Network

In research for the application of Bayesian Network, information has been collected on pattern recognition[11, 12], image classification[13] and automatic image segmentation[14]. There are different kinds of algorithms for Bayesian network classifier as well, such as Naïve Bayes, Tree Augmented Naïve Bayes and Selective Naïve Bayesian Network[15], etc.

The Bayesian network classifier is based on a simple Bayes theorem[16]:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \qquad \text{(eq. 4.4.1)}$$

A and B are events, and P(A), P(B) are the probability of event A and B without regard to each other. P(A|B) is a conditional probability describing the probability of event A given that B is true, also called a posterior. P(B|A) is the likelihood term describing the probability of observing B given that A is true. During the classification, a set of training data is required to train the Bayesian Network. For each class, the network treats each feature as random variables and obtains a certain distribution from the data.

The classification is then done by[15]:

$$c^* = \arg\max_{j=1\ldots m} P(c_j) \prod_{i=1}^{n} P(A_i = a_i | c_j) \qquad \text{(eq. 4.4.2)}$$

This means that given new data, for each feature, the classifier will calculate the probability that it belongs to class j (j=1…m) and assign the class that has the highest posterior for this data.

There is an open source MATLAB toolbox for Bayesian network[17] which can be utilized for building a classifier. In our recent work, a simple simulation was done to test the robustness of the method.

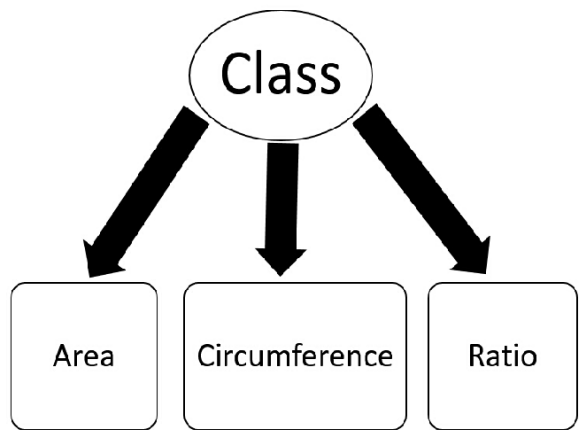
Imagine that we decided to extract three features from each image, namely area of the damage, the circumference of the damage and the aspect ratio of the damage section. Then a

pool of 300 sets of this type of data was generated randomly by MATLAB. The data has 3 kinds of damage, 100 sets each, and the value for each feature of each class follows the same normal distribution. In Figure 4.4.1. (a), column A is the class number, while B, C and D are the area, circumference and aspect ratio respectively. As shown in Figure 4.4.1. (b), the Bayesian network uses a simple network with features that have direct relation with class.

From the 300 data sets, we randomly selected 250 sets as the training group and the other 50 sets as the test group. Comparing the result from the test group with the actual class label of the data sets, we found that the classifier had a near perfect prediction with an accuracy close to 100%.

	A	B	C	D
1	1	0.6900	3.8341	0.1800
2	1	1.8307	3.2882	0.5567
3	1	1.7394	4.7111	0.3692
4	1	2.0296	4.4803	0.4530
5	1	2.8175	3.8549	0.7309
⋮	⋮	⋮	⋮	⋮
101	2	1.3435	10.4699	0.1283
102	2	1.4128	9.9582	0.1419
103	2	0.9822	10.3133	0.0952
104	2	1.4345	7.6099	0.1885
105	2	0.4970	11.2668	0.0441
⋮	⋮	⋮	⋮	⋮
201	3	27.6736	120.5221	0.2296
202	3	28.7156	126.3752	0.2272
203	3	24.5057	119.0940	0.2058
204	3	30.3437	116.3882	0.2607
205	3	30.3360	128.3063	0.2364
⋮	⋮	⋮	⋮	⋮

(a)



(b)

Figure 4.4.1. (a) The random generated data; (b) The Bayesian network

The perfect prediction may be because the network is treating the random variables as if they follow a normal distribution such that the data that was generated coincides with the algorithm that classifies it. Further research in the algorithm and principle of the network is to be done in the second phase of the project.

A maximum entropy network is proposed to combine with the Bayesian network in order to add a constraint to the value of certain features. In the maximum entropy network, the posterior is given as:

$$p(\theta) \propto \mu(\theta)\mu(x'|\theta)e^{\beta \cdot g(\theta)} \qquad \text{(eq. 4.4.3)}$$

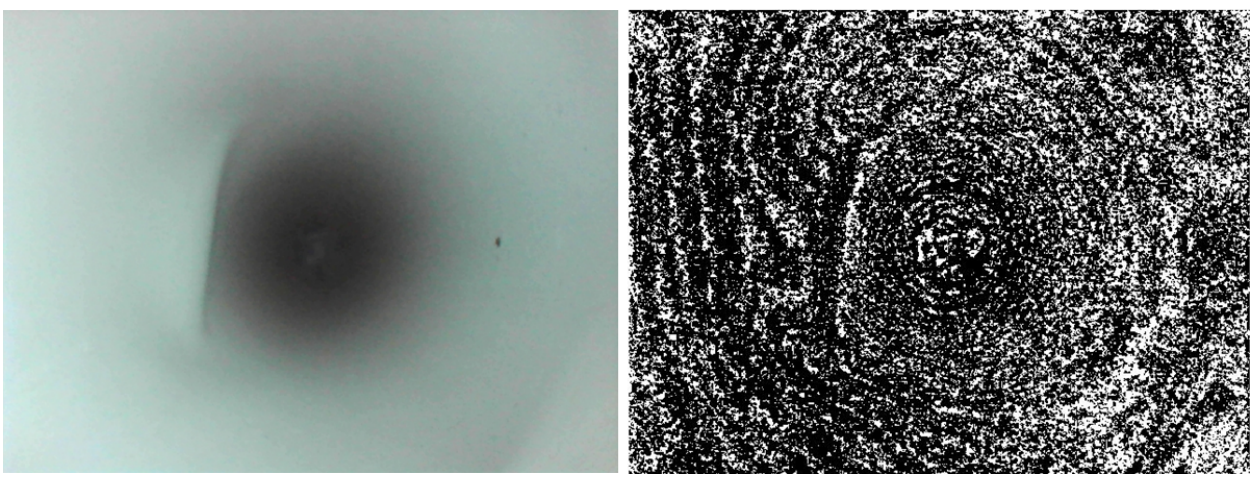
where $\mu(\theta)$ is the prior, $\mu(x'|\theta)$ is the likelihood function and $e^{\beta \cdot g(\theta)}$ is the exponential term introduced by the constraint[8]. Note that when there is no constraint available (i.e. β is zero), the posterior of the maximum entropy network is identical to the Bayesian method. In the proposed project, the maximum entropy network is to be combined with the Bayesian network to form a dynamic network, in an attempt to improve the behavior of the classifier.

4.4.3 Image analysis

Since the Bayesian Network is to be used to do the classification, feature extraction from the raw image is needed. In addition, the images taken with our endoscope camera are noisy; noise deduction and image processing techniques become a necessity for the project.

The first idea is by comparing an image of damaged pipe with that of an intact pipe. The perfect pipe image can be subtracted from the damaged one to show the damage. In a MATLAB help document[19], a structural element was used to create a morphological opening. The opening operation has the effect of removing objects that cannot completely contain the structuring element. Figure 4.4.2 is our trial in this method. As we can see, the result is less than ideal for this application. The reason for this shortcoming is because the image is too noisy or the size of the structural element needs to be optimized.

The pros for this method are that the processing speed is fast, and it is easy to use since it is all MATLAB built in functions. While the drawback is that the size of the morphological opening needs to be adjust for every image with different size of damage, which is not quite possible since we will be dealing with a large amount of data.

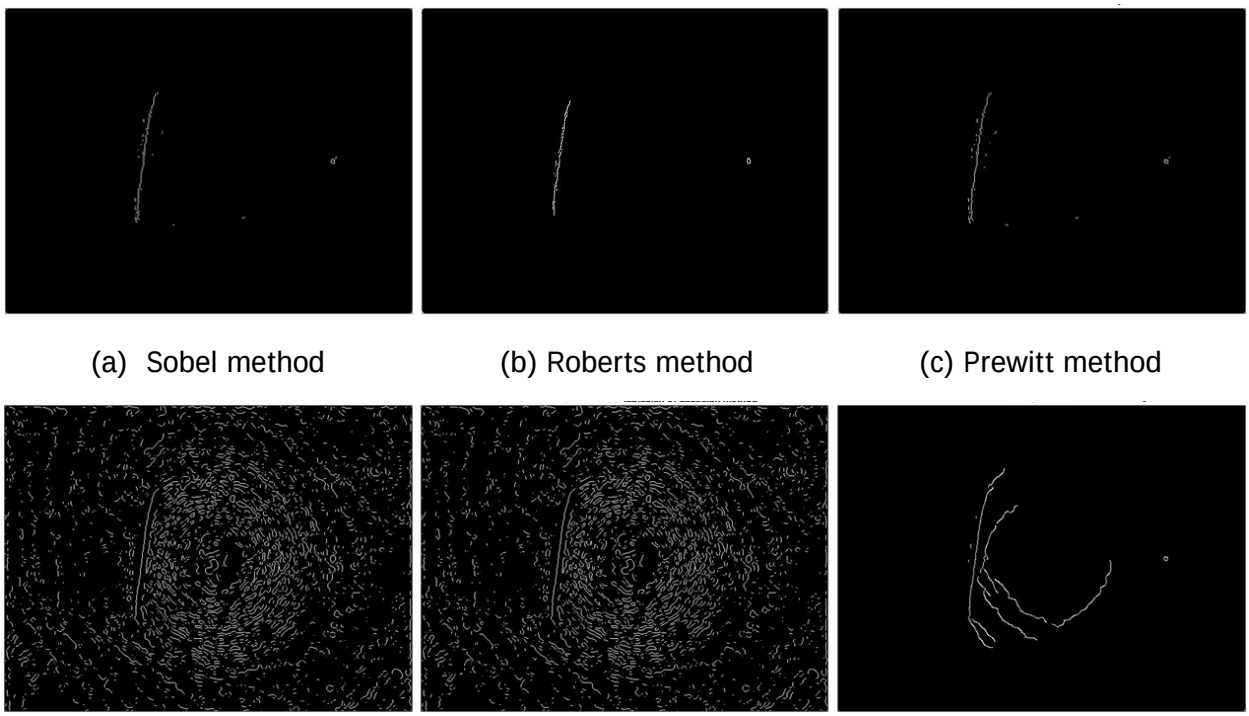


(a) Original image (b) Result after subtracted background

Figure 4.4.2 Illustration of the imaging analysis for feature extraction

Another method called edge detection was researched. Edge detection is an image processing technique for finding the boundaries of objects. It is frequently used in image segmentation and data extraction[20]. There are several methods for edge detection, however they either find places where the first derivative of the image is larger than a certain threshold or the second derivative has a zero crossing[21]. For digital images, edge detection was done by the convolution of an operator with the original image matrix. The process functions to approximate the derivatives in the image. The operator often comes in pair, having one for horizontal and one for vertical.

There are six built-in methods in MATLAB, namely Canny, Laplacian of Gaussian, Prewitt, Roberts, Sobel and the zero-cross method[22, 23]. Canny, Prewitt, Roberts and Sobel method look for the maximum gradient (first derivative), and the Laplacian of Gaussian and zero-crossing methods finds edges by looking for zero-crossings (second derivative equals zero). As the result shown in Figure 4.4.3, the first category seemed to have a better recognition for the damage that we specified. Stains and dust are still unwanted information in the image.



(a) Sobel method (b) Roberts method (c) Prewitt method

(d) Zero-cross method (e) Laplacian of Gaussian method (f) Canny method

Figure 4.4.3. Results of edge detection of image samples

The desired result for the image processing is to be able to isolate the damage so that features can be extracted from the damage; this method is to be explored more to achieve this goal. A proposed method is to use a laser pattern to achieve higher contrast, however the obstacles would be to make the laser coaxial with the camera, while still fitting into the pipe.

4.4.4 Hardware design for imaging stabilizer

The noise reduction for imaging is critical for the successful damage feature extraction. The above discussion focuses on the software/algorithms. Hardware (i.e., using an imaging stabilizer in pipe) is also helpful to reduce the noise due to vibrations and endoscope movement. Very few studies have been performed to design a device that can hold a camera in a pipe with a diameter less than 2 inches. A preliminary conceptual design for imaging stabilizer is shown in Figure 4.4.4. The endoscope camera can be put in the middle tube, while the spring provides tension for the slider on the main body so that the outer arms are pushed out in the

radial direction. In this way, the wheels will have pressure against the pipe wall so that the camera can be held steadily in the pipe.

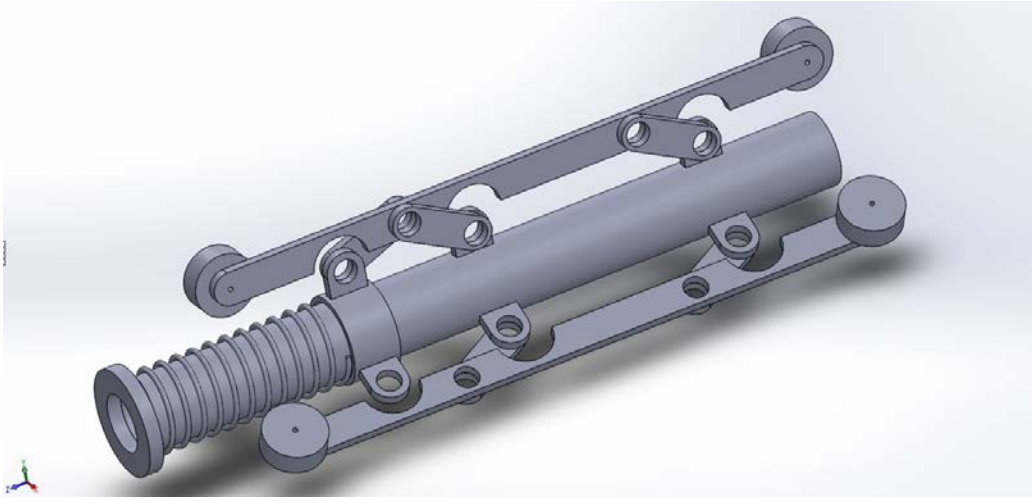
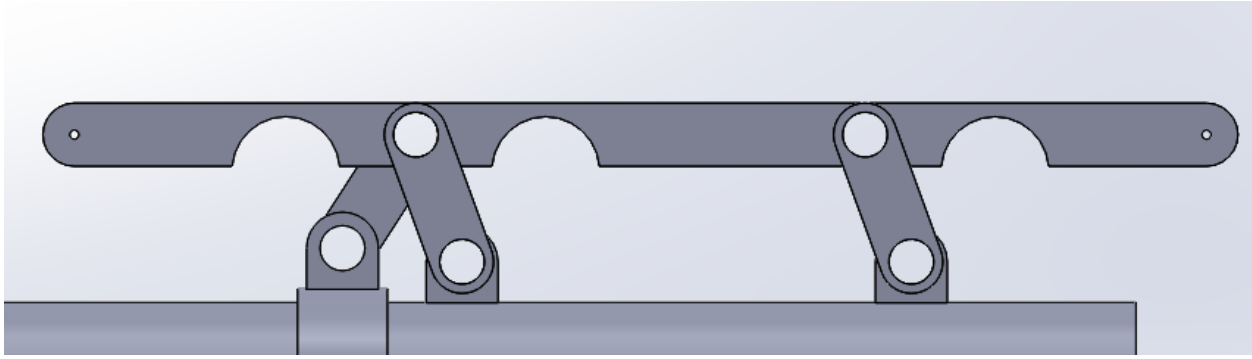
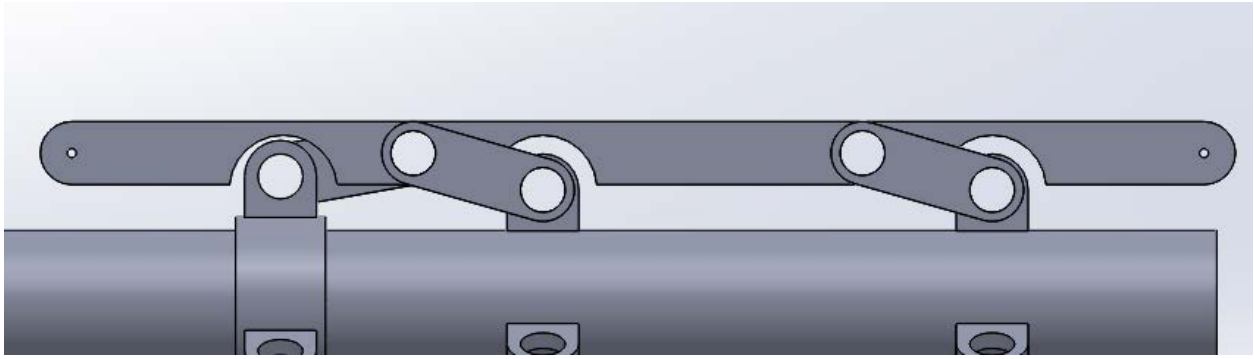


Figure 4.4.4: CAD model for the design prototype

The notches on the outer arms are designed to make it possible for the device to fit in a small pipe with the ability to traverse obstacles such as large dents or a squeeze-off damage. When extended (Figure 4.4.5. (a)), the three arms will hold the camera in the center of the pipe's cross section. When it encounters an obstacle, the device can squeeze in and crawl over the obstacle. This is the limiting factor for the smallest pipe diameter with which this device can be inserted.



(a) The outer arm in extended form



(b) The outer arm in compressed form

Figure 4.4.5: A schematic for how would the design work

Since we already have a 3D printer in our lab, the next step is to utilize it to print out all parts and assemble then with bolts and nuts of the right size. A demonstration is planned to for the next reporting period.

4.4.5 GPU programming

In reality, the pipeline infrastructure will be thousands of miles long, which means an enormous amount of data to be processed. Hence, GPU programming is proposed to handle the image processing. GPU is the graphic processing unit typically used for computer graphics computation. It is principally designed for accelerating image creation for output to a display[24, 25]. GPUs have a massive parallel array of integer and floating-point processors and well as high-speed memory with thousands of cores, which is ideal for highly parallelized computational works. Though it can achieve better performance with highly parallelized computing, some of the drawbacks are the size of the memory on board the device and the transfer speed between system memory and device memory[26].

MATLAB provides a Parallel Computing Toolbox that supports GPU programming with a CUDA-enabled NVIDIA GPU, and many of MATLAB's built in function can be applied. CUDA stands for Compute Unified Device Architecture[27]. It is a parallel computing platform and API (application programming interface) model created by NVIDIA and it allows developers to use GPUs for general purpose processing. The CUDA platform is a software layer that gives direct access to the GPU's virtual instruction set and parallel computational elements[28]. Applications running on NVIDIA graphics processors enjoy superior performance per dollar and performance per watt than implementations built exclusively on traditional central processing technologies[29].

In preparation for this project, an EVGA GeForce GTX 980Ti was purchased and successfully installed (Figure 4.4.6.). The model is 06G-P4-4995-KR. It has 2816 CUDA cores and is clocked at 1102 MHz. The device memory is 6GB GDDR5 with a 384-bit interface. The potential computational power of the graphic card is to be further explored.

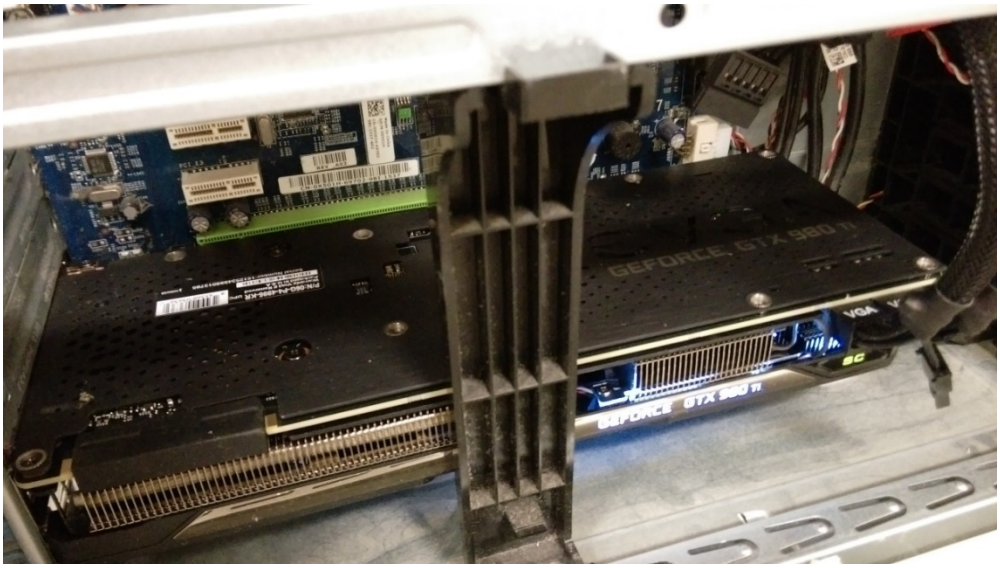


Figure 4.4.6: The GTX 980 Ti Graphic card on board.

The progress that has been made includes: research in the Bayesian network and its application as a classifier, applying maximum entropy methods, exploration of algorithms for image analysis, prototyping hardware for stabilizing the camera and a preparing a high-end GPU for GPU parallel programming.

The future work will be to continue research in the field of image analysis, and work closely with CU Denver for the algorithm for image processing and feature extraction. The hardware design for stabilization device will be 3D printed and assembled in the near future. Coding techniques in parallel computing and GPU programming will be learned and utilized for our application.

4.5 Additional Summary

References

1. Boenisch A., *Advanced subsea pipeline inspection: an overview of flied experience and benefits of subsea pipeline*, Pipelines International, 2013.
2. Barnwell C., and Baker M., *LiDar for Terrain Mapping on the Alaska Pipeline Corridor*, White paper, Alaska Natural Gas Transportation Projects Report, 2009.
3. Duran, O, Althoefer, K, Seneviratne, L.D., *Automated pipe defect detection and categorization using camera/laser-based profiler and artificial neural network*, Automation Science and Engineering, IEEE Transcations on, 2007. **4**(1): p. 118-126.
4. Lipkin K., *Laser Technologies: Taking the Guesswork out of pipeline condition assessment*, Trenchless Technology, 2012.
5. Schmalz, C., et al., *An endoscopic 3D scanner based on structured light.* Medical Image Analysis, 2012. **16**(5): p. 1063-1072.

6. Jimenez L., and Landgrebe D., Supervised classification in high dimensional space: Geometrical, statistical, and asymptotical properties of multivariate data, IEEE Trans. Syst., Man, Cybern, C. 1998, **28**(1): p. 39-54.
7. Guan C., Hassebrook, L. G., and Lau, D. L., Composite structured light pattern for three-dimensional video, Optics Express, 2003, **11**(5): p. 406-417.
8. 8. R. Gonzalez, R. Woods and S. Eddins, Digital Image processing using MATLAB. Upper Saddle River, N.J.: Pearson Prentice Hall, 2004.
9. 9. Schmalz, C.; Angelopoulou, E., "A graph-based approach for robust single-shot structured light," in Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on , vol., no., pp.80-87, 13-18 June 2010
10. A. Fitzgibbon, "Robust registration of 2D and 3D point sets", Image and Vision Computing, vol. 21, no. 13-14, pp. 1145-1153, 2003
11. Jayech, K., M.A. Mahjoub, and N. Ghanmi. *Application of Bayesian networks for pattern recognition: Character recognition case*. in *Sciences of Electronics, Technologies of Information and Telecommunications (SETIT), 2012 6th International Conference on*. 2012. IEEE.
12. Likforman-Sulem, L. and M. Sigelle, *Recognition of degraded characters using dynamic Bayesian networks.* Pattern Recognition, 2008. **41**(10): p. 3092-3103.
13. Jayech, K. and M.A. Mahjoub, *Clustering and Bayesian network for image of faces classification.* arXiv preprint arXiv:1204.1679, 2012.
14. Zhang, L. and Q. Ji, *A Bayesian network model for automatic and interactive image segmentation.* Image Processing, IEEE Transactions on, 2011. **20**(9): p. 2582-2593.
15. Friedman, N., D. Geiger, and M. Goldszmidt, *Bayesian network classifiers.* Machine learning, 1997. **29**(2-3): p. 131-163.
16. Kendall, M., A. Stuart, and J. Ord, *Vol. 1: Distribution theory*1994: London [etc.]: Arnold [etc.].
17. Murphy, K. *Bayes Net Toolbox for Matlab*. 1997-2002; Available from: https://code.google.com/p/bnt/.
18. Guan, X., R. Jha, and Y. Liu, *Probabilistic fatigue damage prognosis using maximum entropy approach.* Journal of intelligent manufacturing, 2012. **23**(2): p. 163-171.
19. *Basic Image Enhancement and Analysis Techniques*. Available from: http://www.mathworks.com/help/images/image-enhancement-and-analysis.html.
20. Umbaugh, S.E., *Digital image processing and analysis: human and computer vision applications with CVIPtools*2010: CRC press.
21. Marr, D. and E. Hildreth, *Theory of edge detection.* Proceedings of the Royal Society of London B: Biological Sciences, 1980. **207**(1167): p. 187-217.
22. Canny, J., *A computational approach to edge detection.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1986(6): p. 679-698.
23. Parker, J.R., *Algorithms for image processing and computer vision*2010: John Wiley & Sons.
24. Fung, J. and S. Mann. *Computer vision signal processing on graphics processing units*. in *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*. 2004. IEEE.
25. Fung, J., F. Tang, and S. Mann. *Mediated reality using computer graphics hardware for computer vision*. in *Wearable Computers, 2002.(ISWC 2002). Proceedings. Sixth International Symposium on*. 2002. IEEE.
26. Corporation, N. *CUDA C Best Practices Guide*. 2010; Available from: http://docs.nvidia.com/cuda/cuda-c-best-practices-guide/#axzz3ts5snRvz.

27. Anand Lal Shimpi, D.W. *"NVIDIA's GeForce 8800 (G80): GPUs Re-architected for DirectX 10"*. 2006 [cited 2006 November 8,]; Available from: http://www.anandtech.com/show/2116/8.
28. Abi-Chahla, F. *Nvidia's CUDA: The End of the CPU?* 2008 [cited 2008 June 18,]; Available from: http://www.tomshardware.com/reviews/nvidia-cuda-gpu,1954.html.
29. Sanders, J. and E. Kandrot, *CUDA by example: an introduction to general-purpose GPU programming*2010: Addison-Wesley Professional.
30. Ansgar S, Thomas F, Carsten S, and Steffen S. 2009. F--a model of events based on the foundational ontology dolce+DnS ultralight. In *Proceedings of the fifth international conference on Knowledge capture* (K-CAP '09). ACM, New York, NY, USA, 137-144.

5. Schedule & Issues

Figure 2 shows the time schedule as of 12/31/2015. No issues were encountered this Quarter.

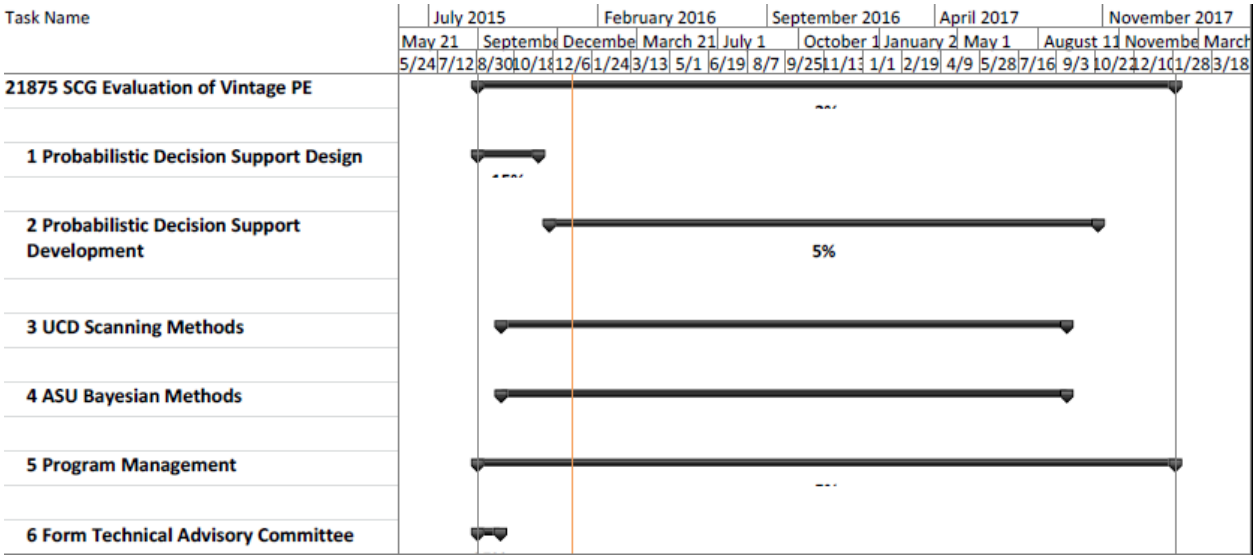


Figure 2 - Time schedule by tasks, as of December 31, 2015

[END OF REPORT]