An Evaluation Framework for Automated Pavement Distress Identification and Quantification Applications

Nima Kargah-Ostadi, Ph.D., P.E. (Corresponding Author)
Research Engineer, Fugro Roadware, Inc.
8613 Cross Park Drive
Austin, TX 78754
Tel: (512) 977-1800
Email: nkargah-ostadi@fugro.com

Abdenour Nazef, P.E.
Pavement Systems Engineer, Florida Department of Transportation
5007 Northeast 39th Avenue
Gainesville, FL 32609
Tel: (352) 955-6322
Email: Abdenour.Nazef@dot.state.fl.us

Jerome Daleiden, P.E.
Director of Pavement Engineering, Fugro Roadware, Inc.
8613 Cross Park Drive
Austin, TX 78754
Tel: (512) 977-1800
Email: jdaleiden@fugro.com

Yuxiao Zhou, M.Sc.
Graduate Engineer, Fugro Roadware, Inc.
8613 Cross Park Drive
Austin, TX 78754
Tel: (512) 977-1800
Email: mzhou@fugro.com

Submitted for Presentation at the 96th Annual Meeting of the Transportation Research Board and Publication in the Transportation Research Record

Revised: November 11, 2016

Word Count
Abstract: 251
Text: 4,950
References: 284
Figures: (7*250)
Tables: (1*250)
Total: 7,485
ABSTRACT
Tracking pavement deterioration types and extent is critical to maintain road networks in a serviceable condition. The prevailing methods for obtaining pavement condition data include manual and semi-automated surveys, which are time-consuming and involve significant human intervention. In response, extensive research has been performed in automating the process for more efficient, objective and repeatable distress evaluations. This paper highlights the preliminary results from an effort sponsored by the Florida Department of Transportation to develop and implement an automated software for identification and quantification of pavement surface cracking distresses.

A technical framework was developed for systematic evaluation of available automated technologies in contrast to manual methods. Pertinent performance measures were identified to evaluate the accuracy, precision, repeatability, reproducibility, and efficiency of various methods. This framework was implemented to determine the gaps in effectiveness of automated applications, design corresponding solutions, and gauge reliability expectations accordingly.

The evaluation follows two main steps: 1) comparison of the cumulative quantities of various distress types found in the manual versus automated surveys, and 2) verification of the automatically detected distresses against reference crack maps generated through a semi-automated process of manually rating the collected images. While the overall comparison of distress quantities provides an indication of strengths and weaknesses of the evaluated algorithm, the distress by distress verification of software performance is used to identify design solutions to address the indicated weaknesses. The guidelines in this systematic framework can be modified with context-sensitive considerations to be applicable to other highway agencies seeking a transition towards automated applications.
INTRODUCTION
Knowledge of pavement deterioration type and extent is critical to cost-effectively maintain road networks in a good and safe condition. The traditional methods for obtaining pavement condition data include manual and semi-automated surveys, which involve significant human intervention and have proven to be time-consuming given the extensive length of road networks. The successful development and implementation of automated crack detection software will result in the following potential advantages to data collection and pavement management initiatives by every roadway agency:

- Increased safety of data collection staff
- Increased efficiency and productivity of network-level evaluations
- Enhanced objectivity of crack rating (identifying type and severity)
- Increased accuracy and precision in measuring crack extent, identifying crack location, and providing summary statistics
- Better pavement management decisions and improved rehabilitations activities

To support the decision-making process within the Florida Department of Transportation (FDOT) Pavement Management System (PMS), the State Materials Office (SMO) performs annual pavement condition surveys (PCS) to evaluate the condition of the State roadway network. The annual PCS data is also used to meet the reporting requirements for the FHWA Highway Performance Monitoring System (HPMS). Currently, the annual PCS is conducted through a manual windshield survey by FDOT distress raters.

FDOT owns and operates a self-contained van equipped with automated and integrated data collection systems (1) equipped with a two-dimensional (2D) Laser Road Imaging System (LRIS). The LRIS digital images have been primarily used to analyze cracks of rigid pavements in areas where manual surveys are difficult and/or unsafe to conduct, especially on urban high-speed roadways. Pavement surface distresses are determined one image at a time by a rater at an office work station using a point and trace manual method. This approach is very inefficient especially for long projects and network level evaluation.

The FDOT has awarded a project to Fugro Roadware to develop and implement automated crack detection applications for rigid and flexible pavements based on the 2D digital images collected using the LRIS. This paper highlights the results from the preliminary evaluation tasks of the project, which were instrumental in identifying the needs and the pertinent design of an automated crack detection and quantification algorithm for rigid pavements. From here on, and for the purpose of this paper, pavement distress refers specifically to pavement cracking distresses.

PAVEMENT DISTRESS EVALUATION METHODS
Pavement distress surveys are typically conducted using one of three approaches: manual, semi-automated, or fully automated. In manual distress surveys, raters perform a visual survey of pavement distresses through either walking on/along the pavement surface, or a windshield survey from a slow moving vehicle.

In semi-automated condition evaluations, raters use a point and trace manual method at an office work station to evaluate the type, extent, and severity of pavement surface distresses one image at a time. This approach is very time consuming especially for long projects and network level evaluation. Currently, most State Highway Agencies (SHA) are using this semi-automated approach (2, 3), which involves significant human intervention.

Fully automated distress evaluations are conducted using image processing and pattern recognition software for distress identification and quantification. Raters conduct quality
assurance testing of the software functionality, and perform quality control of the distress ratings output by the software. The following general steps are typically followed to conceptualize an algorithm for automated pavement distress surveys:

1) Initial Processing: These initial image processing steps improve and enhance image quality for better crack detection.

2) Pre-filtering: These algorithms help identify the approximate regions where no cracks are evident in the image. This helps save processing time by limiting the computationally expensive detection operations to be limited only to regions where there is a high chance of occurrence of cracks. It also helps reduce the number of false positives.

3) Crack Detection: After pre-processing and pre-filtering, the digital image is run through a crack detection routine that uses one or a combination of different image processing techniques like intensity-threshold-based, edge detection, transform-based, seed-based, and/or machine-based methods.

4) Crack Quantification: Once the cracks are detected, they are quantified to calculate geometric properties like crack nodes, length, width, area, density, bounding box, axis of orientations, photometric properties like texture parameters, average intensity, intensity variations, and longitudinal/transverse linear and spatial location referencing.

5) Crack Classification: Once the cracks are identified and quantified, a pattern recognition algorithm can be applied to classify the cracks based on the established protocols. Axis of orientation can be used to classify Transverse and Longitudinal cracks; crack density can be used to classify pattern cracking (such as map cracking or fatigue); width, density, and position of cracks can be used to classify cracks by severity, proximity to other detected cracks or joints and also assist with classification.

A range of national, international, and state specific standards have been created to facilitate automation, such as the American Association of State Highway and Transportation Officials (AASHTO) provisional protocol (PP) number 67 (4) and the United Kingdom’s Surface Condition Assessment of the National Network of Roads (SCANNER) specifications (5). However, these standards are a work in progress and still in various stages of adoption. In these protocols, distresses are identified according to their geometric characteristics. This is unlike the manual and semi-automated distress procedures, where distresses are identified according to the crack generation mechanism. For example, in lieu of identifying fatigue/alligator cracking, the extent and severity of “pattern” cracking is measured within each of the five identified zones. Also instead of discrete severity levels, the continuous average crack width is recorded to account for the intensity of surface distresses. In this manner, SHA have the flexibility to change the severity thresholds and apply this criterion to past data. The AASHTO provisional protocol (PP) number 68 (6) calls for some basic requirements in terms of image quality.

The NCHRP survey conducted in 2004 and updated in 2008 showed that across North America, 44 out of 65 (68%) surveyed agencies use automated pavement data collection (2, 3). The majority of these agencies (30 out of 44) are still reluctant to use a fully automated approach for distress identification. In an effort to select an appropriate automated approach, several agencies have conducted evaluation and comparison experiments on available software (7).

Unlike the advanced data collection technology available in the pavement industry, the pattern recognition software is in need of further enhancements to accurately detect and classify the various types of pavement surface distresses. Also, the lack of a true reference makes it extremely difficult if not impossible to accurately establish the true effectiveness of such systems. The data and images collected today provide an opportunity for reprocessing in future years, when
future crack detection algorithms have larger capability and faster processing speed with increased computing power. This means that collecting the data today may have additional long-term analysis potential.

Only a few highway agencies have implemented a fully automated crack detection system for network-level data collection. From the existing experience, a systematic quality management process seems to be the central consideration for a successful implementation (8). Such a quality management system requires quantitative quality investigations and corresponding acceptable (and context-sensitive) thresholds to achieve a consistent level of quality. These components are determined through an evaluation of the automated technologies in terms of accuracy, precision or reliability, efficiency, and the benefits to the agency.

STUDY OBJECTIVES
It was determined early into the literature review task that there was a need for a systematic framework for benchmarking and evaluating various available distress survey methodologies. This framework is essential in determining the gaps in automated survey algorithms and devising corresponding context-sensitive solutions. This paper serves the following objectives:

1. Explain the devised framework for evaluation of existing methodologies
2. Provide the results of the evaluation of existing distress survey methods using the devised framework in the context of FDOT needs
3. Explain the context-sensitive gap analysis and software design recommendations

FRAMEWORK FOR EVALUATION OF AUTOMATED PAVEMENT DISTRESS APPLICATIONS
The flowchart in FIGURE 1 illustrates the devised evaluation plan for conceptualization and design of the automated crack identification and quantification algorithm for rigid pavements. The key tasks and the corresponding results of this systematic framework are detailed in the subsequent sections of this paper.
VALIDATION OF COLLECTED IMAGE QUALITY

Before evaluation of the available distress survey methodologies, there was a need to validate the optimum quality of the images collected with the FDOT LRIS and provide guidelines for future routine validation. To conduct such an evaluation, an existing FDOT site with measured surface features and targets was used as the control. The following factors were investigated in this validation process, as they might affect the subsequent crack detection:

1. General Image Properties: minimum resolution, appropriate exposure, dynamic range, white balance
2. Image Issues: alignment of the established longitudinal/transverse control lines, prevalent streaks in LRIS images
3. Image Feature Capturing (optical distortion): crack length and width in longitudinal, transverse, and diagonal orientations, signal-to-noise ratio
4. LRIS Hardware: distance measuring accuracy, latitude-longitude accuracy, LRIS platform stability
5. Environmental Effects: performance at various lighting conditions, temperature, humidity, and wind, performance at varying speeds

In addition, the image quality requirements of AASHTO PP68 (6) were evaluated in the validation process. These image quality descriptions are extremely dependent on the crack detection algorithms that are used to evaluate whether the extent of cracks of specific width could be detected, or whether false positives are avoided. Therefore, these quality descriptors cannot be considered independently from the crack detection algorithm used by each agency. It is recommended that each agency uses their corresponding crack detection software to evaluate whether their collected images in combination with their software meet the requirements of the existing provisional protocol. Ideally however, the protocol needs to be modified to provide standard image quality requirements independent of the type of applied crack detection algorithm.
The result of this benchmark testing on the control site determined that the FDOT vehicle collected LRIS images that were of adequate quality to be used for automated crack detection and quantification, given appropriate precautions are practiced. The pertinent precautions were documented as a hardware calibration protocol to be used by FDOT staff to ensure long-term image quality and consistency.

DISTRESS SURVEY PROTOCOL
Florida DOT’s Rigid Pavement Condition Survey Handbook includes the following distress types and severities that are related to cracking (9):

1. Transverse Cracking (count), Light-Moderate-Severe
2. Longitudinal Cracking (count), Light-Moderate-Severe
3. Spalling (linear feet), Moderate-Severe
4. Corner Cracking (count), Light-Moderate-Severe
5. Shattered Slabs (count), Moderate-Severe

The focus of this study was mainly the cracking distresses. Guidelines are provided in the FDOT handbook for establishing extent and severity of each distress type. These guidelines are reviewed with the designated FDOT raters each year to confirm they are current on their understanding of the guidelines and expectations. In order to improve consistency among the FDOT raters and to further clarify the protocol considerations of this experiment, a two-day distress raters’ workshop was organized, which included classroom as well as field training exercises. Based on the meeting discussions and the results of the field exercise, several notes were recorded for consideration in the quantification process required for software development. In addition, FDOT staff are considering these notes to potentially include them for further clarity of the handbook.

REPRESENTATIVE TEST SECTIONS
To conduct an evaluation of the existing methods, FDOT engineers identified a set of 12 representative test sections each at least a standard evaluation length of 0.1-mile and contain several of the jointed concrete pavement distresses in them, making sure that all the crack distress types and severities are incorporated in at least a couple of the sections selected. In the selection of the representative test sections, the following major factors were considered to the extent possible:

- Existence of a variety of cracking distresses
- Existence of different severity levels for each crack type
- Lighting conditions (with or without shades) and angle (going into or out of the sun)

Due to the project limitations, the pavement surface texture and concrete tinning factors were not considered in the selection of the representative sections. Pavement images of these 12 test sections were collected during three (3) repeat runs with the FDOT MPSV (2D LRIS).

SUCCESS METRICS
The three principal success metrics of any process are effectiveness, efficiency, and reliability. In the context of automated distress identification, effectiveness can be expressed in terms of accuracy of the crack detection software when compared to a reference baseline. Accuracy is a qualitative term referring to whether there is agreement between a measurement made on an object and its true (target or reference) value. Bias is a quantitative term describing the difference (or error) between the average of measurements made on the same object and its true value.
While systematic errors identified in the bias can be calibrated out, this is not the case for random errors. The average results may be quite comparable, but individual results can deviate significantly. Reliability of automated distress surveys is often expressed in terms of precision, which is a qualitative term describing the degree of repeatability of a measurement value. Variance and standard deviation of error are quantitative estimates of precision. In addition to precision, reproducibility and repeatability are other important measures of reliability.

The success metrics for automated condition evaluations were considered for two aspects of the process:

1. Comparison of the overall cumulative quantities of various distress types found in the existing manual windshield survey, manual rating of the collected images (semi-automated survey), and a readily available automated software
2. Verification of the automatically detected distresses against the reference crack maps generated through a manual evaluation of the collected images (semi-automated survey)

REFERENCE RATING

Multiple research efforts in the past have introduced various methods to establish reference values or “ground truth” for pavement surface defects (8). The types of reference values mostly used include:

1. Manual distress identification: In this method, professional (trained and experienced) distress raters identify the surface distresses on a set of pavement sections that are deemed representative of conditions across a network.
2. Semi-automated detection: This method is similar to the manual method, but the professional raters use the images collected by monitoring vehicle to identify the distresses.
3. Artificially fabricated distress: In this method, cracks are designed and fissured into steel plates or cut into an existing pavement surface, so the precise dimensions of the defects are known. This method is mostly used to resolve image distortion issues caused by the camera lens or the image sensor.

Each method has its own advantages and limitations. The manual method simulates the actual distress identification process that has been used by many SHA for a long period; however, there is a low degree of agreement among different professional raters which renders the reference as a highly variable measure. The semi-automated option might be superior to the manual method for establishing reference values, because the collected images are available for multiple raters to view in an office environment with less distraction from field traffic. However, some of the low severity cracks that are at initial stages of development might not be visible from the collected images. Increasing the number of raters could result in a more reliable reference.

After much deliberation, it was decided to use a semi-automated approach, in which images from one run of the 12 test sections were rated by one engineer and then completely reviewed and corrected by 2 other engineers. The two advantages to this method is that first the rating was controlled by two additional raters, and second that the raters were not FDOT experienced raters and therefore this reference can be used as an unbiased reference to evaluate all of the three rating methods. FDOT raters and the software engineer did not have access to this reference rating.
COMPARISON RESULTS ON OVERALL DISTRESS QUANTITIES

Based on the overall cumulative amount of each distress among different test sections and multiple runs, the following success metrics were used to compare different rating methods:

- **Average error (bias)** was used to represent *accuracy* or effectiveness of each method. Accuracy can only be quantified with respect to a reference value.
- **Average standard deviation of error** among 12 sections was used to represent *precision* of each method. This reliability metric is also dependent on the reference values.
- **Reproducibility** of manual and semi-automated methods was evaluated by the agreement among the 3 raters in rating the same test sections. This agreement was expressed by 100 minus the coefficient of variation in the total amount of each distress type among the 3 raters. This reliability metric is independent of the reference values.
- **Repeatability** of the automated algorithm was estimated by 100 minus the coefficient of variation in the total amount of each distress type among the 3 runs. This reliability metric is also independent of the reference values.
- In addition, the number of outliers among the 3 raters/runs was represented by the number of section ratings that were more than 1 standard deviation away from the average of the 3 raters/runs. This number is also a reliability metric independent of the reference values.
- **Efficiency** among the three methods was expressed by the estimated time required for each survey method.

FIGURE 2 compares the accuracy of the manual field survey, semi-automated rating, and the automated algorithm compared to the reference survey for each distress type (all severities combined) in each section. Accuracy is calculated as 100 minus the absolute value of bias (average error in percentage). Errors were calculated as the difference between each value and the reference, normalized to the reference and expressed in percentage. It should be noted that the absolute bias of the manual and semi-automated methods in rating spalling was above 100 percent.

The automated algorithm was used for classifying transverse and longitudinal cracks. The considered automated routine was not capable of rating corner cracks and shattered slabs. The selected algorithm was capable of identifying transverse and longitudinal joints, but it was not equipped to distinguish concrete slabs from one another. Although the algorithm estimated crack lengths, it did not produce a crack count per slab (as it was required by the FDOT method). Furthermore, the evaluated software could not detect spalling using the existing images. It is expected that depth (3D) data is required for the detection of spalling at joint or crack edges.
The results indicate that the automated routine is relatively more successful in detecting and identifying the length of transverse cracks (accuracy of 83%) compared to longitudinal cracks (accuracy of 71%). In fact, the accuracy of the automated routine in terms of the transverse cracks (83%) is comparable to the accuracy of the manual field surveys (75%), but it is lower compared to semi-automated surveys (93%). The reason for lower accuracy of the automated rating of longitudinal cracks is mainly because of the longitudinal joints or lane stripes that were falsely classified as cracks and therefore increasing the number of cracks (false positive bias).

FIGURE 3 shows a comparison of precision among the methods as calculated by the standard deviation of error among 12 test sections. The automated rating has lower precision compared to the other rating methods. This is indicated by the higher variation in the rating error among the 12 test sections. This is because the automated algorithm performed much better for some sections compared to the others. In test sections where any of the following defects exists, the automated algorithm had significantly larger errors causing the standard deviation of error among sections to be larger:

- Skewed transverse joints
- Skewed longitudinal joints
- Sawed-in longitudinal joints that are not perfectly straight lines
- Pavement marking (white or black stripes within the lane)
- Traffic counters, and other surface scratch marks
FIGURE 3. Precision Comparison of Different Methods (spalling variation was above 100%)

FIGURE 4 compares the reproducibility of the manual and semi-automated methods to the repeatability of the automated software method using the agreement among multiple raters/runs in rating each test section. The automated algorithm is showing comparable repeatability in identifying transverse cracking among multiple runs, when compared to the reproducibility among multiple raters in the other rating methods. However, the repeatability of the automated method is lower than other methods in rating longitudinal cracking. It should be noted that this is with different sets of images on each different run. This is because of the differences in levels of shades in multiple images of the same pavement section. If the same set of images are used, then there is zero variability among multiple runs of the software.

A review of FIGURE 2, FIGURE 3, and FIGURE 4 reveals that the semi-automated results show higher accuracy (lower bias) and higher precision (lower standard deviation among sections) compared to the manual field surveys in terms of distress extent. However, the semi-automated
rating method yields lower agreement (higher standard deviation) among raters compared to field surveys. These results suggest that with further training and practice, the semi-automated results will probably yield higher reproducibility in addition to the currently higher accuracy and precision compared to the manual field surveys.

FIGURE 5 shows the percentages of outlier ratings as a measure of reproducibility of the manual and semi-automated surveys, and repeatability of the automated rating. Outliers are defined as ratings that were more than one (1) standard deviation away from the average of the three (3) raters/runs. The results indicate that for transverse cracking, the semi-automated rating method yields a better agreement with the least number of outliers among raters. For longitudinal cracking however, there is more agreement among the raters in the manual survey but the least number of outliers is in the semi-automated method.

![Percentage of Ratings More Than 1 Standard Deviation Away from Average](image)

**FIGURE 5. Percentage of “Outlier” Ratings, A measure of Reproducibility/Repeatability**

Efficiency of the manual, semi-automated, and automated methods can be evaluated by the amount of time required to conduct each survey type. FDOT raters provided an estimate for the amount of time to conduct a windshield survey of the 12 test sections. Based on this crude estimate, the manual rating speed in doing windshield surveys while driving on the shoulder is about 1 to 3 miles per hour, depending on the amount of distresses present. It should be noted that for approximately 98% of the concrete pavements, they can drive on the shoulder. The software was used to extract the amount of time that each rater had spent on each test section to conduct a semi-automated survey. The automated detection, classification, and rating takes about 20 seconds per image frame. FIGURE 6 shows a comparison between the semi-automated and automated survey methods in terms of survey speed in miles per hour for each test section and each rater/run.
FIGURE 6. Comparison of the Efficiency (Speed) of Survey Methods

The automated software has an average speed of 0.68 mph (based on the selected detection settings), which is more than twice the efficiency of the semi-automated method. If the detection settings are changed, the automated software could be as fast as 2.7 mph, but that would result in lower accuracy. The manual windshield survey is about twice faster than the automated survey. However, the automated survey does not require human intervention while running and only some QC is required after survey completion. Therefore, more computing power is needed rather than human intervention. In addition, the automated and semi-automated survey methods have the significant advantage of eliminating safety concerns for the raters who are driving or walking on highway shoulder.

VERIFICATION OF AUTOMATICALLY DETECTED DISTRESSES

While the overall comparison of the quantities of each distress type provide an indication of the strengths and weaknesses of the automated methodology, there is a need for a distress by distress verification of the software performance to identify the reasons behind the previously indicated weaknesses. Based on the reference semi-automated crack maps, the following metrics were evaluated on a distress by distress basis:

- **True Positives**: correctly detected cracks (or distress)
- **False Positives**: detected cracks that don’t exist in the reference survey
- **False Negatives**: Missed cracks
- **Distress Validity (or Accuracy)**: an indicator calculated as the ratio of the correctly detected cracks (true positives) to the total detected cracks (true positives and false positives). This statistic indicates the percentage of the detected distress that was actually present in the reference survey, thereby expressing the validity of the distress detected by algorithms.
- **Distress Sensitivity (or Recall)**: a parameter calculated as the ratio of the correctly detected cracks to the total actual cracks existing on the pavement surface (true positives and false negatives). This statistic represents the percentage of the distress in
the reference survey that was detected by the automated method, thereby expressing
the sensitivity of the algorithms to existing distress.

- **Distress Classification Performance**: a measure of the number of correctly classified
cracks (according to the reference survey), divided by the number of correctly detected
cracks (true positives). This statistic indicates the percentage of the detected distress
that is correctly classified by the automated algorithm.

These metrics were evaluated on 24 sample image frames (2 image frames randomly
selected from each of the 12 test sections on run number 1) and the results can be found in FIGURE
7. The results clearly indicate that once a crack is appropriately detected (within the true positives),
the algorithm is doing a good job in terms of assigning the correct distress type (note the blue bars
matching the hatched area).

![FIGURE 7. True Positives, False Positives, and False Negatives in Comparison to Reference Values](image.png)

However, the overall length of false positives and false negatives (missed cracks) are not
comparable. A relatively reasonable balance between the two extremes of aggressive detection
(false positives) and missing cracks (false negatives) would be an indication that appropriate
detection settings have been used in the software. There needs to be an effort in the future
development initiatives to maintain this balance and perhaps improve it. Details of the approach
for selecting these settings have been documented elsewhere (10).

Based on the results in FIGURE 7, the distress verification metrics were calculated. The
results indicate that on average, 78% of the automatically detected distresses actually existed on
the reference survey (78% accuracy), 84% of the distresses on the reference survey were detected
via the automated algorithm (84% recall), and 86% of the detected distresses were correctly classified into longitudinal or transverse cracking or joints.

**GAP ANALYSIS**

TABLE 1 summarizes the identified gaps and corresponding recommended solutions.

<table>
<thead>
<tr>
<th>Number</th>
<th>Category</th>
<th>Gap</th>
<th>Recommended Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Human Random Errors</td>
<td>High variation of rating results among test sections</td>
<td>N/A</td>
</tr>
<tr>
<td>2</td>
<td>Human Systematic Errors</td>
<td>High bias (average error) and high variation of rating results among multiple raters</td>
<td>Review and/or revise distress protocols</td>
</tr>
<tr>
<td>3</td>
<td>Software Systematic Errors</td>
<td>High bias in longitudinal cracking amount (high number of false positives)</td>
<td>Joint detection plugin</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>High variation of error among multiple test sections</td>
<td>Joint detection plugin</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>Issue with crack counts</td>
<td>Improve crack grouping</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>Not rating corner cracks</td>
<td>Corner crack plugin</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>Not rating shattered slabs</td>
<td>Shattered slab plugin</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>Issue with crack width determination after image filtering and therefore issues with severity rating</td>
<td>Change pixel intensity threshold for measuring crack width</td>
</tr>
<tr>
<td>9</td>
<td>Hardware Limitations</td>
<td>distresses such as spalling or patching cannot be detected without 3D data</td>
<td>Evaluate 3D data</td>
</tr>
</tbody>
</table>

**DESIGN RECOMMENDATIONS**

Based on the gap analysis, the following development efforts are recommended for the next task of this project. The algorithm logic design is briefly explained for each development effort which is aimed at improving the capabilities of the existing software:

1) Transverse and longitudinal joint detection plugin
   
a. First, the image is downsized (the size of downsized image is 0.25*0.25 of original image) and the downsized image is divided into a matrix of cells. The size of each cell is 8 by 8 pixels. Then, the darkest pixel in each cell is detected and the information (i.e. intensity and coordinate) of the darkest pixel is assigned to that cell. The existence of a joint cell is then validated by comparing its intensity with the intensity of adjacent cells using a contrast index. If the cell intensity is lower than adjacent cells (darker cell), then the cell is a candidate joint cell.

   b. A narrow moving window at 0 or 90 degrees (for transverse or longitudinal joints) is used to scan the whole candidate joint cell map, which only contains joint cells. The number of cells that fall into this window is counted. If the number of cells is larger than a threshold number (for example, 1/10 of width or height of downsized image), then the candidate joint pixel detection result is probably a true positive and the centerline of the window is considered as joint coordinates. Otherwise, the detection result is a false positive and would be excluded from further analysis.
c. Some crack cells (cells that belongs to a pavement crack) could be source of false positives. A second narrower window is used to filter out this kind of error. The gap between the candidate joint cells in the narrow window is calculated and if this gap is higher than a specified threshold, then this is a meandering crack (going in and out of the narrow window) and not a straight joint.

d. Finally, the selected joint coordinate (x coordinate for longitudinal joint and y coordinate for transverse joint) is recorded, and a line is drawn on the image. The coordinate information is also recorded in the SQL database for future assigning of cracks to specific slabs.

2) **Plugin to improve crack grouping and count per slab**

   a. Group cracks based on the types of cracks (longitudinal or transverse)

   b. Sort the longitudinal cracks based on minimum value of y coordinate (minY), and sort the transverse cracks based on minimum value of x coordinate (minX)

   c. Connect the crack pair C1(startPoint1(sx1, sy1), endPoint1 (ex1, ey1)) and C2(startPoint2(sx2, sy2), endPoint2 (ex2, ey2)), if and only if one of the following exists:

      i. Both of the cracks belong to the category of longitudinal cracks, where sy1<sy2 AND ey1<sy2 AND |sx2-ex1|/|sy2-ey1|<=1 (the trajectory between the cracks is also longitudinal) AND Euclidean distance between endPoint1 and startPoint2 is smaller than a threshold value (for example, 500 pixels) OR

      ii. Both of the cracks belong to the category of transverse cracks, where sx1<sx2 AND ex1<sx2 AND |sx2-ex1|/|sy2-ey1|>=1 (the trajectory between the cracks is also transverse) AND Euclidean distance between endPoint1 and startPoint2 is smaller than a threshold value (for example, 500 pixels)

   d. Extract the corner points (vertices) of the slab and count the cracks in each slab based on the relationship of crack ends and slab vertices:

      i. Assume crack startPoint (Xsp, Ysp) and endpoint (Xep, Yep) coordinates, and the slab’s upper left point is (X1, Y1) and lower right point is (X2, Y2).

      ii. If (Xsp is between X1 and X2) AND (Ysp is between Y1 and Y2), then we consider the startPoint (Xsp, Ysp) is inside the slab. If (Xep is between X1 and X2) AND (Yep is between Y1 and Y2), then we consider the endpoint (Xep, Yep) is inside the slab.

      iii. If any crack startPoint or endPoint is inside the slab, the crack is regarded as a ‘crack in this slab’.

3) **Plugin to classify and rate corner cracks**

   a. extract the corner cracks based on the relationship between endpoints of crack and slab edges (or joints); A corner crack should intersect with both transverse and longitudinal joints, and this point of intersection should be more than 1 foot apart from the slab corner on both the longitudinal and transverse joints. Based on its location, we classify corner cracks into four types: upper left, upper right, bottom left and bottom right.

   b. rate the corner crack based on its width.

4) **Plugin to classify and rate shattered slabs**
a. count the number of regions bounded by the 'cracks inside the slab' and joints. If the number is greater than 4, the slab is counted as shattered slab. Note that the grouping plugin should have satisfactory performance before this plugin can successfully classify shattered slabs.

b. rate the shattered slab based on width of crack.

SUMMARY AND CONCLUSIONS
This paper described the developed systematic framework to compare various pavement distress survey methodologies. Pertinent performance measures were identified to evaluate the accuracy, precision, repeatability, reproducibility, and efficiency of various methods.

The results of implementing this framework to compare manual, semi-automated, and automated methods were presented for rigid pavement distress surveys according to FDOT distress protocol. The results of this evaluation framework were instrumental in identifying the gaps within the existing automated software and determining corresponding solutions to address them.

While the overall comparison of distress quantities provided an indication of strengths and weaknesses of the evaluated algorithm, the distress by distress verification of software performance was used to identify design solutions to address the indicated weaknesses. The guidelines in this systematic framework can be modified with context-sensitive considerations to be applicable to other highway agencies seeking a transition towards automated applications.

REFERENCES
